

## THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Optique et Radiofréquences**

Arrêté ministériel : 7 août 2006

Présentée par

**Julie ROULLARD**

Thèse dirigée par **Bernard FLECHET** et  
codirigée par **Stéphane CAPRARO**

préparée au sein du **Laboratoire IMEP-LAHC, site de Chambéry**  
dans **l'École Doctorale EEATS**

# Analyse et optimisation des performances électriques des réseaux d'interconnexions et des composants passifs dans les empilements 3D de circuits intégrés

Thèse soutenue publiquement le **15 décembre 2011**,  
devant le jury composé de :

**Mme. Valérie MADRANGEAS**

Professeur à l'Université de Limoges,

Président

**Mme. Emmanuelle BOURDEL**

Maître de Conférences à l'ENSEA Cergy Pontoise,

Rapporteur

**Mr. Eric LHEURETTE**

Professeur à l'Université de Lille,

Rapporteur

**Mr. Gilles POUPON**

Ingénieur au CEA/LETI, MINATEC Grenoble,

Invité

**Mr. Alexis FARCY**

Ingénieur Docteur à ST Microelectronics, Crolles,

Invité

**Mr. Stéphane CAPRARO**

Maître de Conférences à l'Université de Saint Etienne,

Co-directeur de thèse

**Mr. Bernard FLECHET**

Professeur à l'Université de Savoie,

Directeur de thèse









# Remerciement

---

Je tiens tout particulièrement à remercier mon directeur de thèse, le professeur Bernard Fléchet, pour sa grande patience, sa diplomatie et son enthousiasme, qui m'ont permis de mener mes travaux à leurs termes.

Etant à l'origine de cette thèse, je lui dois énormément et garderait en tête des instants et des endroits incroyables, comme Miami Beach, le restaurant au dessus du César Palace, le Crasy Elephant ou encore le Hammam de Tanger... Bref des souvenirs nombreux et merveilleux qui resteront inoubliables.

Toujours présent et réceptif, il est un peu le superman du laboratoire.

Je voudrais également remercier mon Co-directeur de thèse, le maître de conférences Stéphane Capraro. Il a su m'accompagner, m'écouter et me conseiller durant ces trois années. Toujours là pour moi, il a été un pilier indispensable à l'avancement de mes travaux. Bien entendu sa mauvaise foi aux cartes n'entachera en rien le merveilleux souvenir de notre collaboration.

Je remercie les membres du jury qui ont accepté d'évaluer ce travail :

Je remercie Madame Valérie Madrangeas, Professeur à l'Université de Limoges, qui m'a fait l'honneur de présider ce jury.

Je remercie Madame Emmanuelle Bourdel, Maître de conférences à l'ENSEA de Cergy Pontoise et Monsieur Eric Lheurette, Professeur à l'Université de Lille, d'avoir accepté la lourde tâche de rapporteur sur ce travail. Je tiens à les remercier pour leur lecture attentive du manuscrit ainsi que pour leurs remarques pertinentes.

Je remercie Monsieur Gilles Poupon, Ingénieur au CEA- LETI, pour ses remarques constructives, son enthousiasme et d'avoir pris le temps de participer à ce jury.

Je remercie également Monsieur Alexis Farcy, Ingénieur à ST Microelectronics, pour sa contribution à mes travaux. Bon nombre de mes résultats sont le fruit de nos conversations. Merci pour ta bonne humeur et tes conseils pertinents.

Je remercie également Léa Di-Cioccio du CEA LETI, et son équipe, qui m'ont permis d'aborder les problématiques de collage de façon approprié.

Je tiens à exprimer toute ma gratitude à Cédric Bermond et Thierry Lacrevez. Toujours partant pour apporter leurs aides et leurs savoirs dès qu'un doctorant a un problème. Ils sont un peu comme le ying et le yang de l'équipe chipi. Lorsque l'un affirme : « Arrête tout, c'est de la m---- » l'autre répond : « Attends mais c'est génial ». Je reste persuadé que ce doux mélange permet au laboratoire de mener à bien de nombreux projets de qualités.

Le petit couplet suivant est dédié à Sebastien de Rivaz. Plus qu'un collègue c'est devenu un ami. Le fait d'avoir partagé son bureau à mon arrivée m'a apporté énormément de choses. Merci

pour ces verres partagés, ces instants de confiance et de rigolade. Tu es sans doute l'une des personnes qui m'a le plus appris au laboratoire et cette thèse n'aurait pas été la même sans toi.

Je remercie également monsieur Florent le normand, qui par sa nonchalance, son sourire et sa bonne humeur à contribuer de façon majeure à une donner une ambiance agréable et joyeuse au travail. Il a le don pour nous donner le sourire pour la journée. Emilie Hérault, dealeuse de séries, qui comme Florent nous donne la banane pour la journée et même pour la soirée lorsqu'il y a cours de PPG. N'oublie pas mimi j'attends mon invitation pour ton mariage (en costume...). Et Magalie Astic, confidente, toujours prête à donner un coup de mains, mon premier coup de fil du matin était pour toi.

Je garde un souvenir particulier de Ludovic Fourneau qui à chaque conversation me donnait l'impression de plonger dans un monde parallèle. Merci pour cette étrange expérience.

Je tiens également à remercier les autres personnes du LAHC actuelles ou passées, pour leur sympathie et / ou leur accueil : Sylvain, Jonathan, Grégoire, Adriana, Lionel, Thomas, Jérôme, Benjamin B, Benjamin P, Pierre, Yann, Gwénael, Jean-Louis, Frédéric, Jean-François, Elie, Pascal, Lionel, Stéphane, Mélanie, Fernanda.

Je tiens à remercier également mes amis, ceux qui m'ont accompagné, de près ou de loin durant cette thèse : Céline, un soutien moral majeur à qui je souhaite le meilleur pour la fin de ta thèse, Piergil, pour ces nombreuses bouffes entre le Bourget et Lyon. Pupu et Pauline pour l'ambiance des soirées Lyonnaises. Je remercie également ma Lili, le Bourget du lac me rappellera deux choses : cette thèse et nos fous rire lorsque l'on était coloc en fac. Et bien sur les pioupious, en vrac : Adrien, Nico, Gael, Sam, Mat (et oui tu es devenu un pioupiou par alliance)...

J'ai également une pensée pour Ophylie et Maxime. Merci Ophélie, nos bringues au petit savoyard m'ont fait démarrer ces trois années de thèse de façon plus qu'agréable... en effet c'est grâce à vous deux qu'à cette époque j'ai rencontré le père de ma fille.

Bien sûr je ne peux oublier ma famille, mes parents qui m'ont toujours soutenu et accompagné dans mes choix. Merci pour votre confiance. Mes sœurs Sophie et Céline ainsi que mes beaux frères Sébastien et Christophe. Cécé je n'oublierai jamais que tu t'ai toujours soucié de savoir comment j'allais et qu'en cas de besoins, avec Chris vous avez toujours répondu présent. Je remercie également mes neveux et nièces, Andréas, Joshuas et Charlène, qui m'apportent un grand bonheur à chaque fois que je les vois.

Je remercie également les Bellos, Pascale et Laurent, pour leur disponibilité et leur bonne humeur, grâce à eux j'effectue une reconversion en douceur.

Enfin je remercie la personne qui a été au premier plan dans ma vie ces dernières années, mon conjoins Ludovic. Merci pour ta patience et ta gentillesse. Pour ces nombreux allers retours entre Douvaine et le Bourget du Lac et ce par tous les temps. Depuis le début tu m'as supporté et soutenu dans cette aventure.

Maintenant on démarre un nouveau chapitre de notre vie, avec le plus beau des cadeaux, notre fille Léa.







# Sommaire

<i>Introduction générale</i> .....	<i>1</i>
<i>Chapitre 1 : L'intégration 3D</i> .....	<i>9</i>
<b>I POURQUOI INTEGRER EN TROIS DIMENSIONS ?</b> .....	<b>11</b>
I.1 LES LIMITES DE L'INTEGRATION 2D .....	11
I.1.1 La loi de Moore.....	11
I.1.2 Les interconnexions.....	12
I.2 LES SYSTEMES INTEGRES.....	15
I.2.1 Système sur puce (SOC : System On Chip).....	15
I.2.2 Système en boîtier (SIP : System In Package).....	16
I.2.3 Vers l'intégration 3D.....	16
<b>II LES OPTIONS TECHNOLOGIQUES POUR UNE INTEGRATION 3D</b> .....	<b>17</b>
II.1 LE TYPE D'EMPILEMENT .....	18
II.2 L'ORIENTATION DES CIRCUITS .....	18
II.3 L'ALIGNEMENT .....	19
II.4 LE COLLAGE .....	19
II.5 L'AMINCISSEMENT DU SILICIUM .....	21
<b>III LES INTERCONNEXIONS EN INTEGRATION 3D</b> .....	<b>22</b>
III.1 LES INTERCONNEXIONS .....	23
III.1.1 Les interconnexions de la couche de redistribution (RDL).....	24
III.1.2 Les interconnexions de la couche de Back End Of Line (BEOL).....	24
III.1.3 Les via traversant le silicium (TSV) .....	25
III.1.4 Les piliers de cuivre (Copper-Pillar).....	26
III.2 FONCTIONS ASSOCIEES A CES INTERCONNEXIONS. ....	27
III.2.1 Inductance au niveau BEOL.....	28
III.2.2 Inductances aux niveaux RDL. ....	28
III.2.3 TSV en structure coaxiale [Ho, 2008] .....	29
III.2.4 Triple condensateur MIM par TSV [Roozeboom, 2008]. ....	29
III.3 APPLICATIONS ENVISAGEES .....	30
<b>IV MES TRAVAUX DE THESE</b> .....	<b>31</b>

<b>Chapitre 2 : Caractérisation et modélisation des briques élémentaires d'interconnexions dans les empilements 3D de circuits</b>	<b>33</b>
<b>I INTRODUCTION</b>	<b>36</b>
<b>II OUTIL INDISPENSABLE A LA CARACTERISATION : LE DE-EMBEDDING</b>	<b>37</b>
II.1 METHODE DE DE-EMBEDDING GENERIQUE	39
II.1.1 Rappel sur la méthode développée par E.P. Vandamme	39
II.1.2 Méthode de E.P. Vandamme dérivée et adaptée à nos contraintes	40
II.2 METHODE DE DE-EMBEDDING DESTINEES A LA CARACTERISATION DE L'EXPOSANT DE PROPAGATION D'UNE LIGNE DE TRANSMISSION	44
II.2.1 Modèle des lignes de transmission	44
II.2.2 Extraction de l'exposant de propagation $\gamma$	46
II.3 METHODE DE DE-EMBEDDING DESTINEES A LA CARACTERISATION DE L'IMPEDANCE CARACTERISTIQUE D'UNE LIGNE DE TRANSMISSION	47
II.3.1 Méthode basée sur les travaux de D.F. Williams	48
II.3.2 Méthode DUMMC (De-embedding using modification of materials characteristics)	51
II.3.3 Méthode d'épluchage par bloc	57
II.3.4 Méthode de de-embedding de TM. Winkel améliorée	62
II.3.5 Conclusion générale sur les méthodes de de-embedding présentées	68
<b>III CARACTERISATION DES BRIQUES ELEMENTAIRES D'INTERCONNEXIONS DANS UN EMPILEMENT 3D DE PUCES</b>	<b>68</b>
III.1 INTERCONNEXIONS DE TYPE RDL (REDISTRIBUTION LAYER)	69
III.1.1 Description du dispositif de test associé à une interconnexion de type RDL	69
III.1.2 Résultats d'extraction et de modélisation	70
III.1.3 Interconnexion de RDL dans un environnement réel d'intégration 3D	70
III.2 INTERCONNEXION DU BEOL ENFOUIE DANS UN ENVIRONNEMENT 3D	73
III.3 THROUGH SILICON VIA (TSV) HAUTE DENSITÉ	77
III.3.1 Description de la cellule de test mesurée	77
III.3.2 Méthode d'épluchage appliquée pour la caractérisation du TSV	78
III.3.3 Extractions dans les domaines fréquentiel	79
III.4 CU-PILLAR	81
III.5 INDUCTANCES PLANAIRES INTEGREES	81
III.5.1 Description des inductances intégrées sous test	81
III.5.2 Résultats d'extraction	82
<b>IV CONCLUSION</b>	<b>84</b>

## Chapitre 3 : Modèles et analyses des caractéristiques électriques des briques élémentaires d'interconnexions dans les empilements 3D de circuits intégrés ... 87

<b>I</b>	<b>INTRODUCTION.....</b>	<b>90</b>
<b>II</b>	<b>LES INTERCONNEXIONS HORIZONTALES. ....</b>	<b>91</b>
II.1	ETUDE PREDICTIVE SUR LES INTERCONNEXIONS DE RDL.....	91
II.1.1	Description des différents types d'interconnexion de RDL.....	92
II.1.2	Etude dans les domaines fréquentiel et temporel.....	93
II.2	ETUDES PREDICTIVES DES PERFORMANCES DES INTERCONNEXIONS DU BEOL.....	97
II.2.1	Description des interconnexions du BEOL.....	97
II.2.2	Etudes dans les domaines fréquentiel et temporel.....	98
II.3	REGLES DE DESSIN POUR L'OPTIMISATION DES PERFORMANCES DES INTERCONNEXIONS DU BEOL EN	
	INTEGRATION 3D.....	99
II.3.1	Etudes paramétriques sur l'analyse du retard de propagation et la diaphonie pour des	
	interconnexions du BEOL lors du report de puces.....	99
II.3.2	Synthèse des règles de dessin.....	112
<b>III</b>	<b>LES INTERCONNEXIONS VERTICALES. ....</b>	<b>114</b>
III.1	ETUDES PREDICTIVES DES PERFORMANCES DES CU-PILLARS.....	115
III.1.1	Description des différents types de Cu-Pillar.....	115
III.1.2	Analyse fréquentielle et modèles électriques des Cu-Pillars.....	116
III.2	ETUDES PREDICTIVES DES PERFORMANCES DES TSV.....	118
III.2.1	Description des différents types de TSV.....	119
III.2.2	Analyse fréquentielle et modèles des TSV.....	120
III.2.3	Analyse temporelle et performances de propagation des TSV.....	122
<b>IV</b>	<b>INDUCTANCES PLANAIRES .....</b>	<b>122</b>
IV.1	DESCRIPTION DE LA STRUCTURE DE REFERENCE .....	123
IV.1.1	Extraction des paramètres du modèle équivalent de l'inductance.....	124
IV.2	Impact d'un empilement Face to Back sur les performances de l'inductance.....	127
IV.2.1	Influence de l'épaisseur du substrat de silicium de la puce rapportée.....	127
IV.2.2	Influence de la conductivité du silicium de la puce rapportée.....	129
IV.2.3	Influence de l'épaisseur de collage entre les deux puces empilées.....	130
IV.3	ARCHITECTURE FACE TO FACE.....	132
IV.3.1	Influence de l'épaisseur et de la conductivité du substrat de silicium de la puce rapportée.....	132
<b>V</b>	<b>CONCLUSION .....</b>	<b>133</b>

## Chapitre 4 : Intégrité de transmission des signaux sur une chaîne complète d'interconnexions dans les empilements 3D de circuits intégrés ..... 137

<b>I</b>	<b>INTRODUCTION.....</b>	<b>139</b>
<b>II</b>	<b>CONVERSION DU DOMAINE FREQUENTIEL AU DOMAINE TEMPOREL .....</b>	<b>140</b>
<b>III</b>	<b>STRATEGIE DE ROUTAGE .....</b>	<b>141</b>
III.1	DESCRIPTION DES SCENARIOS DE ROUTAGE.....	142
III.2	RESULTATS.....	143
<b>IV</b>	<b>STRATEGIE D'ORIENTATION .....</b>	<b>144</b>
IV.1	DESCRIPTION DES SCENARIOS D'ORIENTATION DES PUCES .....	144
IV.2	RESULTATS.....	145
<b>V</b>	<b>STRATEGIE TECHNOLOGIQUE.....</b>	<b>146</b>
V.1	DESCRIPTION DES SCENARIOS TECHNOLOGIQUES .....	147
V.2	RESULTATS.....	147
V.2.1	<i>Performance en termes de retard.....</i>	<i>147</i>
V.2.2	<i>Influence relative de chaque brique élémentaire d'interconnexion dans une chaîne complète d'interconnexion.....</i>	<i>148</i>
<b>VI</b>	<b>CONCLUSION .....</b>	<b>150</b>

## Chapitre 5 : Evaluation des performances et stratégies d'empilements 3D pour des circuits rapides de type mémoire- processeur (Wide I/O) ..... 153

<b>I</b>	<b>INTRODUCTION.....</b>	<b>156</b>
I.1	SPECIFICATIONS POUR CIRCUIT MEMOIRE-LOGIQUE. ....	158
I.2	LES ALTERNATIVES D'EMPILEMENT. ....	159
<b>II</b>	<b>COMMUNICATION MEMOIRE-LOGIQUE ET LOGIQUE-BGA (AVEC ET SANS INTERPOSER)....</b>	<b>160</b>
II.1	EMPILEMENTS SANS INTERPOSER .....	160
II.1.1	<i>Description des empilements .....</i>	<i>160</i>
II.1.2	<i>Calcul des fréquences maximales de communication.....</i>	<i>160</i>
II.2	L'INTERPOSER : UNE SOLUTION DE PLATEFORME TECHNOLOGIQUE 3D PROMETTEUSE.....	163
II.2.1	<i>Description de l'empilement pour une solution avec Interposer.....</i>	<i>163</i>
II.2.2	<i>Résultats obtenus sur les performances électriques.....</i>	<i>164</i>
II.3	CONCLUSION DES PARTIES 2 ET 3. ....	169

<b>III</b>	<b>COMMUNICATIONS MEMOIRE-PROCESSEUR ET PROCESSEUR-BGA EN FONCTION DES DENSITES D'INTEGRATION (AVEC ET SANS INTERPOSER).....</b>	<b>170</b>
III.1	CAS DE L'EMPILEMENT FACE TO BACK .....	170
III.1.1	Description de l'empilement.....	170
III.1.2	Résultats sur l'empilement Face to Back.....	172
III.2	ROADMAP ET STRATEGIE D'EMPILEMENT POUR LES CIRCUITS ASSOCIANT UNE MEMOIRE A UN PROCESSEUR.....	174
III.2.1	Fréquences maximales en fonction des générations technologiques. ....	175
III.2.2	Stratégies d'empilement selon le type de densité d'intégration.....	176
<b>IV</b>	<b>CONCLUSION .....</b>	<b>176</b>
	<i>Conclusion générale et perspectives .....</i>	<i>179</i>
	<i>Références .....</i>	<i>189</i>
	<i>Annexes .....</i>	<i>197</i>
	<i>Liste des publications .....</i>	<i>211</i>



## Introduction générale

### Contextes actuels

En version mode actuelle et haute couture...

Aujourd'hui l'appellation 3D est partout, elle nous accompagne dans des domaines aussi divers qu'inattendus. Que ce soit au cinéma, à la télévision, par le biais d'une console de jeu, d'une caméra, d'un appareil photo ou d'une imprimante, les effets 3D nous font rêver et génèrent un marché commercial porteur. La technologie n'est pas le seul domaine touché, les visites virtuelles en trois dimensions sont en plein essor. Dorénavant nous avons la possibilité d'admirer en 3D les œuvres exposées au musée du Louvre, d'explorer la vallée de Chamonix ou de visiter virtuellement sa future maison. L'effet marketing provoqué par le sigle 3D est incontestable et ce n'est pas le chanteur Mickey 3D ou la marque de lingerie Dim avec ses sous vêtements pour homme 3D flex, qui soutiendront le contraire (Figure 0-1).

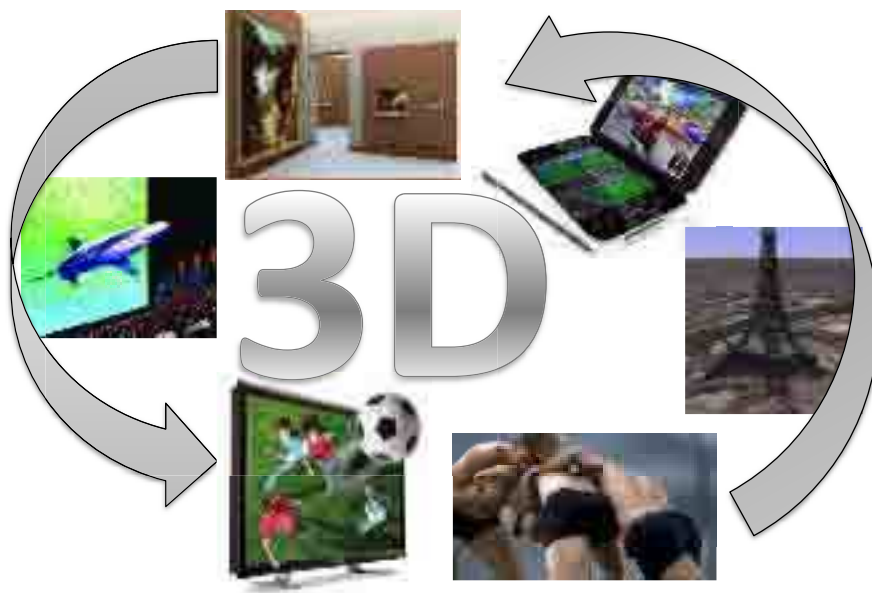


Figure 0-1 : La troisième dimension au quotidien.

En version recherche et développement en microélectronique ...

Dans le domaine de la microélectronique l'intégration 3D de circuits n'est pas qu'un outil de marketing, elle offre des perspectives nouvelles et réellement prometteuses. Depuis plus de cinquante ans, la célèbre loi de Moore a permis de modéliser la tendance à miniaturisation des puces et à l'augmentation des densités d'intégration. Elle est associée à la réduction de la taille des

transistors réalisés en technologie planaire, utilisés dans les processeurs numériques, les mémoires... Cependant, les performances de ces circuits intégrés, longtemps imposées par les caractéristiques des transistors, éléments piliers de la microélectronique moderne, se trouvent aujourd'hui limitées par le réseau d'interconnexions reliant les différents blocs logiques, voire analogiques des puces. En effet, avec la miniaturisation, ces interconnexions supportent des densités de courant de plus en plus élevées, elles présentent des longueurs de plus en plus importantes et doivent véhiculer des signaux dont la fréquence ne cesse de croître. Temps de retard, diaphonie et consommation sont autant de paramètres qui dégradent les performances électriques de ces interconnexions, ainsi que leur fiabilité. Le cahier des charges imposé ou souhaité par l'ITRS [ITRS, 2009] est toujours plus contraignant et de plus en plus difficile à atteindre, et cela se fait au prix d'investissements technologiques considérables.

L'évolution de la microélectronique arrive à un stade où certains signes de ralentissement dans la course vers l'intégration ultime apparaissent. La miniaturisation des composants commence à atteindre ses limites physiques, générant un point d'interrogation de premier ordre quant à la question de l'avenir technologique de la microélectronique actuelle sur silicium. Ces interrogations sont l'occasion de proposer des solutions nouvelles en termes de co-intégration hétérogène et de techniques de traitement et de transport de l'information. L'intégration 3D fait partie de ces concepts prometteurs et technologiquement réalisables à court terme.

L'idée est de prendre séparément des puces assurant chacune une des différentes fonctionnalités nécessaires à l'application visée et de les empiler les unes sur les autres. Cette approche évite des mises en boîtier intermédiaires mais aussi d'imposer sur toute une puce l'utilisation de la technologie la plus avancée nécessaire pour réaliser une seule fonctionnalité contenue dans la puce, la plus exigeante, alors que la réalisation des autres fonctionnalités de cette puce ne le nécessite pas. Il en résulte une augmentation globale du niveau d'intégration, de la fiabilité, des performances électriques, mais aussi une baisse de la consommation électrique ainsi que de l'encombrement [Lèquepeys, 2011]. L'objectif ultime serait d'empiler des puces ayant chacune une fonctionnalité optimisée, c'est à dire réalisée avec la technologie la mieux adaptée afin de construire un circuit complet, ultra performant. Cette intégration 3D hétérogène pourrait, par exemple, permettre d'empiler une puce capteur/antenne RF réalisée sur substrat HR SOI, sur une puce dédiée au de traitement analogique RF sur AsGa, sur une puce CAN sur silicium, sur une puce dédiée à la mémoire et une puce dédiée au processeur de calcul ....

Ceci dit, empiler les puces implique qu'il faut être capable de réaliser un grand nombre d'étapes technologiques très différentes :



1. amincir les substrats. Une fois amincie à quelques dizaines de microns, une puce devient fragile et difficile à manipuler. Il faut donc coller celle-ci sur un autre support plus rigide (poignée) durant les manipulations
2. disposer de techniques de collage inter-puces, permettant le maintien mécanique ainsi que la réalisation de contacts électriques entre elles ou une isolation, selon la zone du circuit considérée.
3. aligner avec une très grande précision les différentes puces
4. disposer de techniques de gravure dans le silicium. Ces trous dans le silicium, permettent, après avoir été remplis par des couches barrières à la diffusion, des couches de métallisation et d'isolation la réalisation des « through silicon-via » (TSV). Ils assurent la connexion électrique entre chaque face de la puce et ainsi le transport des signaux entre les puces empilées ou vers l'extérieur. Ces TSV, interconnexions verticales, jouent un rôle clef dans les technologies d'intégration 3D.

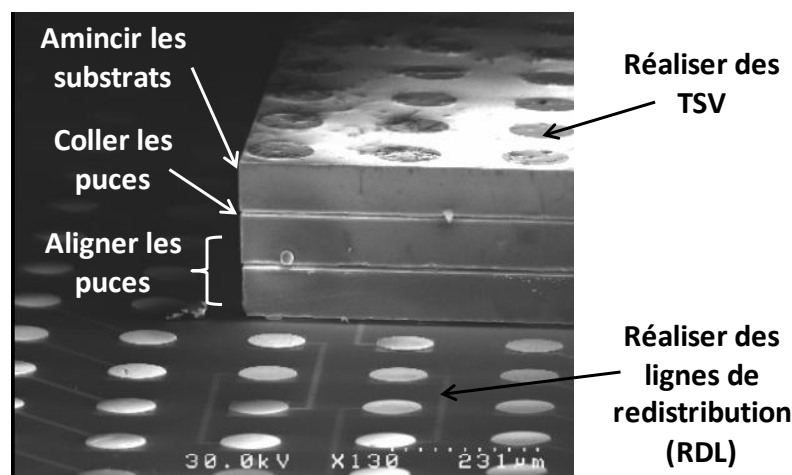


Figure 0-2 : L'intégration 3D et quelques unes de ses étapes technologiques, illustration avec 3 puces empilées (image SEM - IBM Tokyo Research Laboratory)

L'intégration 3D amène une quantité d'options technologiques nouvelles à maîtriser. Néanmoins c'est un axe de recherche qui est, à l'heure actuelle, porté dans le monde entier par une large majorité d'acteurs de la microélectronique.

## Cadre général de ces travaux de doctorat

De nombreuses études sont actuellement menées sur les concepts d'empilement de processeurs logiques et de puces dédiées à la mémoire, pour avoir la mémoire située au-dessus des unités de calcul, le tout en un seul boîtier. Les marchés visés concernent des objets mobiles intégrant des applications embarquées (téléphones, tablettes, jeux vidéo...) sont très exigeants en terme de bande passante, de capacités de mémoire et de faible consommation. Avec ces architectures, les débits d'information entre la mémoire et le processeur peuvent être augmentés (plus de 12 GBy/s sont espérés) et la consommation réduite (moins de 600 mW sont attendus). Finalement l'ordre de

bataille est d'obtenir les meilleures performances en terme de puissance de calcul, avoir la consommation la plus faible, dans l'encombrement le plus réduit et de s'interfacer avec l'environnement grâce à de nouveaux capteurs et actionneurs et de ce fait réaliser des objets toujours plus petits, toujours plus performants, toujours moins gourmands en énergie.

Ceci dit, la réalisation d'une telle application peut être abordée suivant une multitude d'approches technologiques. Quelle est la meilleure orientation des puces lors de leur empilement ? Faut-il utiliser, ou non, une plateforme technologique de type interposer ? Quelle densité d'intégration de TSV répond le mieux au besoin de cette application ? Ce sont autant de questions qui ouvrent un nombre important de possibilités pour la réalisation de circuits 3D. Sans oublier que chaque alternative technologique d'intégration 3D conditionne des architectures d'interconnexions différentes, faisant appel à des briques d'interconnexions nouvelles comme les TSV, les piliers de cuivre ou les lignes de redistribution en face arrière (RDL). L'étude de ces scénarios passe nécessairement par une quantité d'étapes intermédiaires. Cela consiste en l'exacte connaissance de toutes les briques élémentaires d'interconnexion, l'extraction expérimentale de leurs paramètres en fréquence, l'étude de leurs comportements dans le domaine temporel ou encore l'association de ces briques afin d'obtenir des chaînes d'interconnexions.

## **Description des travaux**

Les travaux présentés dans ce manuscrit ont pour objectif de proposer les meilleures pistes à suivre pour optimiser les performances des réseaux d'interconnexion lors de la conception d'un circuit en trois dimensions, et cela pour chaque type d'application visée. Pour parvenir à cet objectif les principales étapes intermédiaires ont été :

- caractériser expérimentalement et modéliser chacune des briques élémentaires d'interconnexions présentes dans un réseau d'interconnexion d'un empilement 3D de puces, sur des spectres de fréquences conformes à ceux des signaux véhiculés dans les différentes applications visées
- développer les outils de simulation des performances électriques des réseaux d'interconnexions 3D, réseaux réalisés en assemblant les différentes briques élémentaires d'interconnexions spécifiques aux technologies d'intégration 3D
- analyser et prédire les performances électriques de ces réseaux d'interconnexions 3D en fonction des différentes alternatives technologiques d'empilement, niveaux d'intégration
- extraire des ces études des règles de conception, c'est-à-dire les meilleures stratégies de routage, d'orientation des puces et de densité d'intégration permettant d'optimiser les performances des réseaux d'interconnexions dans les empilements 3D

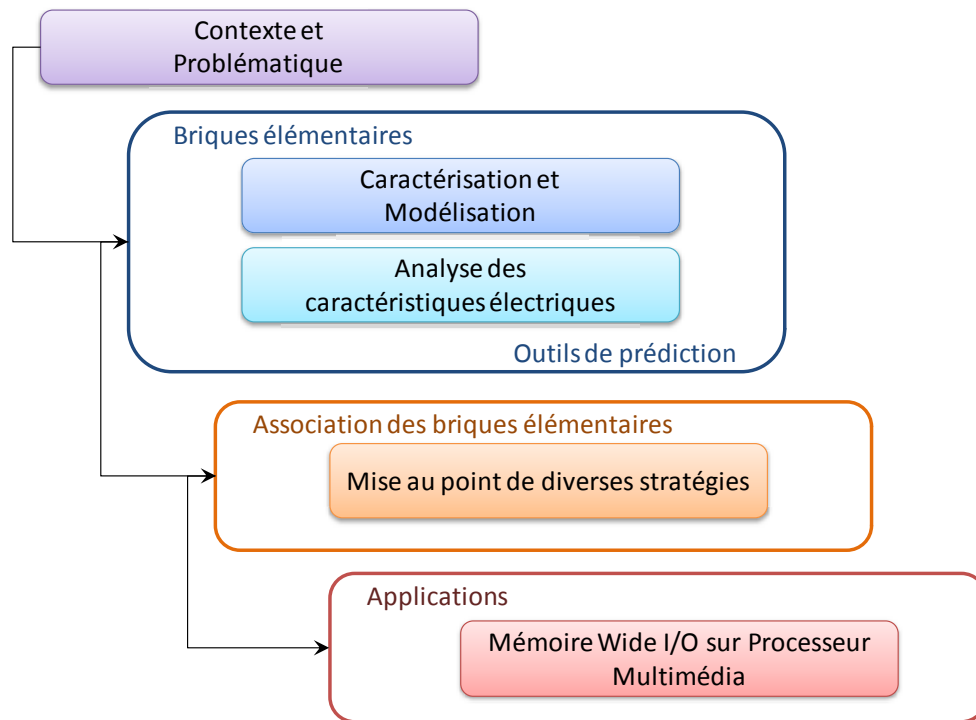


Figure 0-3 : Synthèse des travaux de thèse.

Dans un premier temps, le chapitre 1 introduit la problématique des travaux de cette thèse en se focalisant sur les spécificités de l'intégration en trois dimensions de circuits microélectroniques.

Le contexte dans lequel s'inscrivent ces travaux, les limites de l'intégration 2D ainsi que les options technologiques nécessaires à l'intégration 3D sont exposées. Par la suite, les nouveaux éléments d'interconnexions en intégration 3D sont présentés. Il s'agit des lignes de redistribution (RDL), des interconnexions du Back end of Line (BEOL), les vias traversant le silicium (TSV), les piliers de cuivre (Cu-Pillar). Les inductances intégrées sont également présentées car les performances de ces dernières sont fortement impactées par les technologies d'intégration 3D. La dernière partie de ce chapitre est consacrée à la présentation des objectifs des travaux de thèses, leurs problématiques ainsi qu'à la démarche suivie.

Le chapitre 2 est dédié à la caractérisation expérimentale et à la modélisation électrique de dispositifs passifs intégrés, telles que les interconnexions, dans le domaine des hautes fréquences (typiquement des MHz à quelques dizaines de GHz).

La première partie expose les spécificités des mesures hyperfréquences sur des composants passifs enfouis au sein d'un circuit intégré. Plusieurs procédures d'épluchage ou de « de-embedding » utilisées dans ces travaux sont présentées. Dans la seconde partie les briques élémentaires d'interconnexions présentées dans le chapitre 1 sont caractérisées et modélisées à l'aide des méthodes de de-embedding mises en place. Les résultats expérimentaux obtenus par ces

différentes méthodes d'extraction sont confrontés à ceux obtenus par modélisation électromagnétique à l'aide des solveurs tels que HFSS et Maxwell 2D d'Ansoft, dans le but de les valider et les utiliser pour des études paramétriques dans la suite de nos travaux.

Dans le troisième chapitre, les performances électriques des briques élémentaires d'interconnexions sont étudiées dans les domaines fréquentiel et temporel.

La première partie est consacrée aux interconnexions horizontales de types RDL et du BEOL. Dans un premier temps, les comportements en fréquence ainsi que les retards induits par différentes interconnexions de redistribution sont étudiés. Puis des règles d'optimisation au niveau de leur design sont déterminées pour les interconnexions du BEOL. La seconde partie concerne les interconnexions verticales comme les TSV et les Cu-Pillar. Dans un premier temps les comportements en fréquence de différents piliers de cuivre sont étudiés. Puis une étude fréquentielle permet de dégager des modèles électriques équivalents et de mener une analyse sur les performances en terme de retard des différentes générations technologiques de via traversant le silicium. La dernière partie est consacrée aux inductances intégrées planaires. On analyse l'impact de différentes options technologiques spécifiques à l'intégration 3D sur leur facteur de qualité

Le chapitre 4 aborde, quant à lui, la thématique de l'intégrité des signaux transmis sur une chaîne complète d'interconnexions dans un empilement 3D de puces. Ces travaux s'appuient sur l'ensemble de ceux menés dans les chapitres 2 et 3.

La première partie expose la méthodologie utilisée pour quantifier les performances électriques d'une chaîne d'interconnexions d'un empilement 3D de circuits intégrés. Les parties suivantes mettent en scène trois scénarios différents, afin de déterminer les grands axes à privilégier lors d'une intégration 3D. Trois stratégies seront définies, la première concernant le routage des interconnexions, la seconde portant sur les options d'empilement et la dernière sur la densité d'intégration, guidée par le niveau d'intégration des TSV.

Dans le dernier chapitre nous utilisons tous les travaux précédents pour les mettre en valeur dans le cadre d'applications concrètes d'intégration 3D et sur lesquelles les industriels de la microélectronique sont très friands d'informations étant donnée l'ampleur des marchés économiques visés. Les possibilités d'associer une puce mémoire avec un processeur (cas généralement nommé Wide IO) suivant différentes options empilement, plateformes technologiques et densités d'intégration sont étudiées. Pour chaque configuration les spécificités en termes de fréquences maximales de fonctionnement et d'intégrité des signaux seront analysées (diagrammes de l'œil).

Dans un premier temps nous étudions le cas des empilements Face to Face et Face to Back, sans interposer, afin de déterminer les vitesses de transmission entre puces et avec le monde extérieur (substrat BGA). Puis nous les comparerons aux performances obtenues avec un empilement 3D comportant un interposer. Cette alternative technologique suscite beaucoup d'intérêts car elle offre une grande souplesse de fabrication et des coûts de réalisation moindres.

Afin de généraliser nos études, la dernière partie reprend, sous forme d'abaque, les trois empilements (avec et sans interposer) et présente les spécifications attendues et obtenues de chaque type de communication pour quatre niveaux de densité d'intégration différents. Pour finir des stratégies de conception répondant aux spécificités des applications Wide I/O seront déterminées pour chaque empilement.



# Chapitre 1

## L'intégration 3D

---

Ce chapitre introduit la problématique des travaux de cette thèse en se focalisant sur les spécificités de l'intégration en trois dimensions de circuits microélectroniques.

La première partie a pour objectif de présenter le contexte dans lequel s'inscrivent ces travaux. Les limites de l'intégration 2D sont abordées par le biais d'un bilan sur les performances des réseaux d'interconnexions. En effet, la miniaturisation, toujours plus poussée, contribue à provoquer des temps de retard non négligeables liés aux longueurs et aux fortes résistances et capacités ainsi que des niveaux diaphonie élevés, engendrés par la densité d'interconnexions devenue très forte. Les dernières solutions apportées, consistant à privilégier la densification avec des systèmes sur puce (SOC) ou à privilégier l'hétérogénéité à l'aide de système en boîtier (SIP), ont permis à l'industrie de la microélectronique de suivre, jusqu'à présent, la loi Moore. Cependant les limites de fonctionnement des circuits, en termes de vitesse, de fiabilité et de coût de réalisation, se rapprochent de plus en plus et poussent les concepteurs à de nouvelles innovations, telle que l'intégration 3D.

La seconde partie présente les options technologiques nécessaires à l'intégration 3D. L'ajout d'une dimension pose de nouveaux challenges de fabrication de ces circuits 3D. Il est désormais nécessaire d'étudier et maîtriser plusieurs aspects, absents dans le cas d'une intégration en deux dimensions. Les différents types d'empilement, l'orientation des circuits, les contraintes d'alignement, le collage ainsi que l'amincissement du silicium sont ces nouvelles notions, explicitées dans cette partie.

La troisième partie est focalisée sur les nouvelles structures d'interconnexions en intégration 3D, elle présente ces nouveaux éléments clefs d'interconnexions présents dans l'intégration 3D. Ces éléments sont nommés dans la suite du manuscrit « briques élémentaires d'interconnexions ». Il s'agit des lignes de redistribution (RDL), des interconnexions du Back end of Line (BEOL), les vias traversant le silicium (TSV), les piliers de cuivre (Cu-Pillar). Nous discutons aussi des inductances intégrées car les performances de ces dernières sont fortement impactées par les technologies d'intégration 3D. L'association de ces briques élémentaires d'interconnexions permet de décrire les chaînes d'interconnexion nécessaires à la communication entre un circuit de mémoire et un processeur ou entre un processeur et un BGA et par conséquent les applications visées par cette technologie.

La dernière partie de ce chapitre est consacrée à la présentation des objectifs des travaux de thèses, leurs problématiques ainsi qu'à la démarche suivie.

## Sommaire

<b>I</b>	<b>POURQUOI INTEGRER EN TROIS DIMENSIONS ?</b>	<b>11</b>
I.1	LES LIMITES DE L'INTEGRATION 2D	11
I.1.1	La loi de Moore	11
I.1.2	Les interconnexions	12
I.1.2.a	Hiérarchie des niveaux d'interconnexion	12
I.1.2.b	Temps de retard	13
I.1.2.c	Diaphonie	14
I.2	LES SYSTEMES INTEGRES	15
I.2.1	Système sur puce (SOC : System On Chip)	15
I.2.2	Système en boîtier (SIP : System In Package)	16
I.2.3	Vers l'intégration 3D	16
<b>II</b>	<b>LES OPTIONS TECHNOLOGIQUES POUR UNE INTEGRATION 3D</b>	<b>17</b>
II.1	LE TYPE D'EMPILEMENT	18
II.2	L'ORIENTATION DES CIRCUITS	18
II.3	L'ALIGNEMENT	19
II.4	LE COLLAGE	19
II.4.1	Evolution des techniques de report de puces	20
II.4.2	Le collage direct Cuivre / Cuivre	21
II.5	L'AMINCISSEMENT DU SILICIUM	21
<b>III</b>	<b>LES INTERCONNEXIONS EN INTEGRATION 3D</b>	<b>22</b>
III.1	LES INTERCONNEXIONS	23
III.1.1	Les interconnexions de la couche de redistribution (RDL)	24
III.1.2	Les interconnexions de la couche de Back End Of Line (BEOL)	24
III.1.3	Les via traversant le silicium (TSV)	25
III.1.4	Les piliers de cuivre (Copper-Pillar)	26
III.2	FONCTIONS ASSOCIEES A CES INTERCONNEXIONS	27
III.2.1	Inductance au niveau BEOL	28
III.2.2	Inductances aux niveaux RDL	28
III.2.3	TSV en structure coaxiale [Ho, 2008]	29
III.2.4	Triple condensateur MIM par TSV [Roozeboom, 2008]	29
III.3	APPLICATIONS ENVISAGEES	30
<b>IV</b>	<b>MES TRAVAUX DE THESE</b>	<b>32</b>



# I Pourquoi intégrer en trois dimensions ?

Depuis la commercialisation du premier microprocesseur, par Intel en 1971, l'industrie de la microélectronique suit la loi empirique de Moore : doublement de la densité de transistors tous les deux ans. Cette évolution a été inévitable et a permis de réduire le coût des circuits intégrés tout en améliorant leur vitesse de fonctionnement.

Cependant les interconnexions servant à relier ces transistors se sont naturellement densifiées et leurs longueurs ont augmenté considérablement. A ce jour on compte plus de 10 km de longueur d'interconnexion dans les circuits les plus denses. Pour restreindre les pertes de performances dues à cette densification, l'aluminium a été remplacé par le cuivre, moins résistif et les diélectriques isolants ont été rendus poreux afin de diminuer leur permittivité et par conséquent limiter les retards et la diaphonie entre les interconnexions. Malgré ces efforts, les interconnexions restent l'un des principaux facteurs limitant les performances des circuits intégrés [Meindl, 2001].

## I.1 Les limites de l'intégration 2D

### I.1.1 La loi de Moore

Depuis les années 60, l'industrie de la microélectronique a entamé une course à la miniaturisation. En 1965, Gordon Moore prévoit le doublement du nombre de transistors par puce tous les 18 mois [Moore, 1965]. Cependant en 1975, Moore réévalua sa prédiction en posant que le nombre de transistors par microprocesseurs, sur une puce de silicium, doublera tous les deux ans. Bien qu'il ne s'agisse pas d'une loi physique mais juste d'une extrapolation empirique, cette prédiction s'est révélée étonnamment exacte. Entre 1971 et 2001, la densité des transistors a doublé chaque 1,96 année. En conséquence, les machines électroniques sont devenues de moins en moins coûteuses et de plus en plus puissantes.

En effet la Figure 1-1, représentant le nombre de transistors par processeur en fonction des années, montre qu'un processeur Intel commercialisé en 2010 compte  $10^5$  fois plus de transistors qu'un processeur Intel fabriqué en 1974.

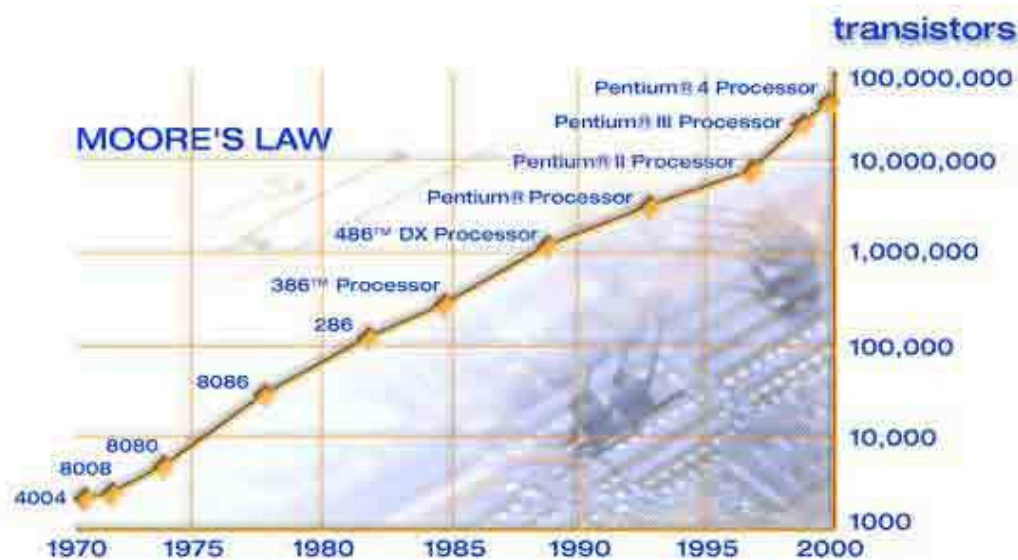


Figure 1-1 : Illustration de la loi de Moore

Cependant ces remarquables avancées technologiques ont un impact non négligeable sur les interconnexions présentes dans le Back End Of Line (BEOL). Dans la partie suivante, les différents niveaux d'interconnexion ainsi que les perturbations (retard, diaphonie) qu'elles engendrent sont explicitées.

## I.1.2 Les interconnexions.

La loi de Moore impacte les dimensions des interconnexions comprises dans les niveaux « locaux » et « semi-globaux ». Les longueurs, épaisseurs et largeurs se voient diminuées à chaque nouvelle génération, ce qui par conséquent fait augmenter leur résistance.

### I.1.2.a Hiérarchie des niveaux d'interconnexion

Un circuit numérique rapide, représenté Figure 1-2 (nœud technologique CMOS 45nm), est constitué de deux blocs fonctionnels différents, nommés Front End Of Line (FEOL) et Back End Of Line (BEOL). Le FEOL regroupe tous les éléments actifs du circuit (porte logique, mémoires...) et est réalisé sur un substrat de silicium. Le BEOL, situé juste au-dessus du FEOL, comprend l'ensemble des composants passifs (inductances, condensateurs, réseaux d'interconnexions...).

Le BEOL de la génération 45 nm est composé de sept niveaux de métallisations dédiés aux lignes d'interconnexions. Le niveau le plus bas (nommé M1) est aussi le plus dense, il est dit « local » et comprend des interconnexions chargées de relier les transistors entre eux. Les quatre niveaux suivants sont appelés « intermédiaires » (ou semi globaux). Les deux niveaux suivants sont quant à eux nommés « globaux » et ils sont connectés au niveau métallique final (Alucap).

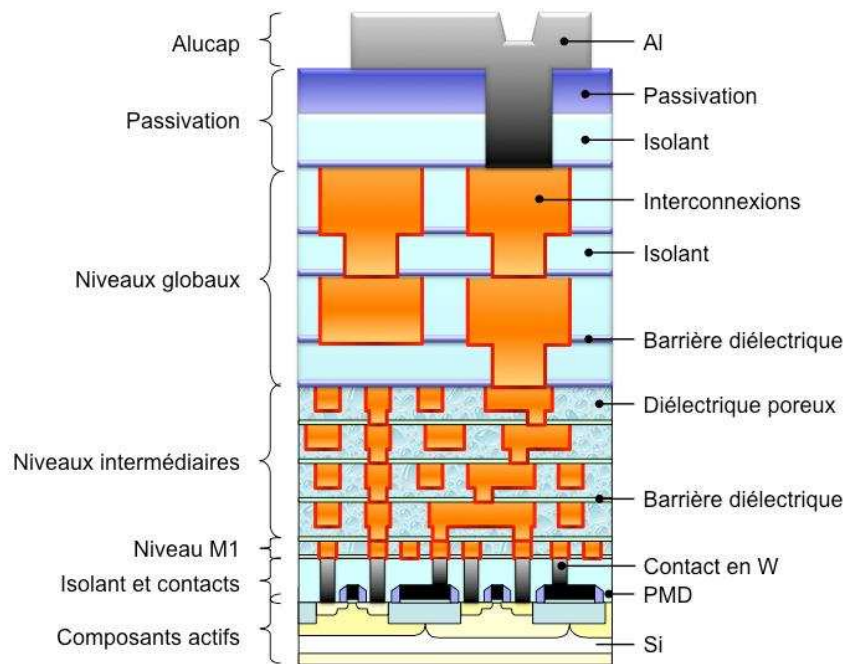


Figure 1-2 : Schématisation d'un empilement FEOL et BEOL d'une technologie CMOS 45 nm.

Entre deux générations successives, la longueur de la grille du transistor est réduite et le nombre de transistors et de niveaux de métal augmentent (Tableau 1-1).

Tableau 1-1 : Evolution du nombre de niveaux de métal et de transistors pour les générations utilisées ces dix dernières années [SIA, 2001].

Années	2004	2007	2010	2013
<b>Génération / Longueur de grille (nm)</b>	90	65	45	32
<b>Nombre de transistors par puce (Millions)</b>	50	145	250	500
<b>Nombre de niveaux de métal</b>	7-8	8-9	9	9-10

### 1.1.2.b Temps de retard

Dans les dernières générations technologiques, les performances en terme de vitesse de commutation des transistors ont été fortement améliorées alors que les performances électriques des interconnexions ont été dégradées [Chen, 2000]. En effet, avec la succession des générations technologiques, la longueur des interconnexions augmente et elles sont de plus en plus proches les unes des autres. L'évolution de ces paramètres implique des comportements résistifs et capacitifs de plus en plus importants et donc des temps de retard et des phénomènes de couplages en constante augmentation.

Le temps de retard correspond au temps qui s'écoule entre les passages à 50 % de la valeur maximale du signal en entrée et en sortie de ligne. Il affecte directement les performances du circuit

en limitant sa vitesse maximale et par conséquent sa fréquence de traitement. Il s'avère également susceptible de générer des décalages temporels entre différents blocs du circuit.

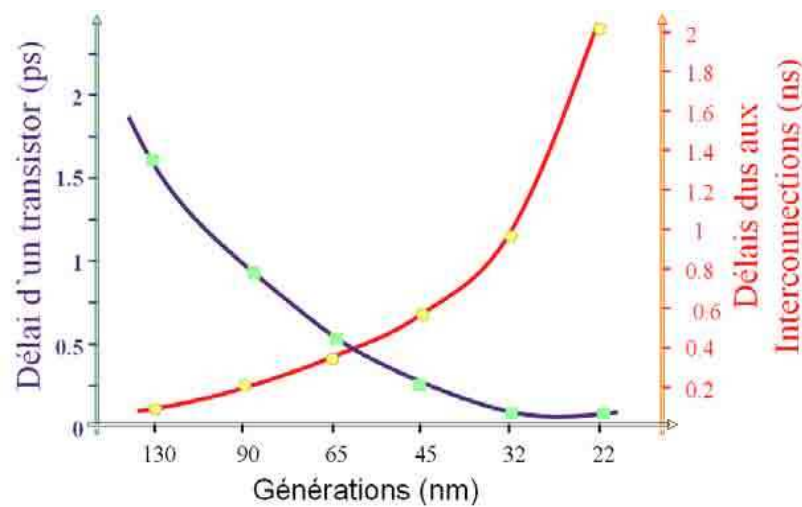


Figure 1-3 : Evolution des délais dans un circuit numérique intégré d'une génération à la suivante.  
[Meindl, 2002]

La Figure 1-3 confirme qu'à partir de la génération de transistors du nœud 180 nm, les temps totaux de fonctionnement des circuits sont majoritairement dus aux temps de propagation dans les interconnexions. Cela est principalement dû au fait qu'à chaque génération de transistors est associé un réseau d'interconnexions toujours plus agressif en terme de dimensionnement et de densification (Tableau 1-2).

Tableau 1-2 : Dimensions des interconnexions intermédiaires.

Années	2004	2007	2010	2013
Génération (nm)	90	65	45	32
Largueur (nm) du niveau intermédiaire	140	100	70	50
Epaisseur (nm) du niveau intermédiaire	240	200	140	100
« Pitch » (nm) lignes intermédiaires	280	200	140	100

### 1.1.2.c Diaphonie

La densification, du fait des géométries agressives (faible espacement entre interconnexions), entraîne un couplage électromagnétique entre les lignes. Il en résulte un phénomène de diaphonie, critère de performance électrique qui doit être pris en compte. Une interconnexion activée par un signal logique, génère alors un signal parasite venant perturber une interconnexion voisine dite victime (principe illustré Figure 1-4).

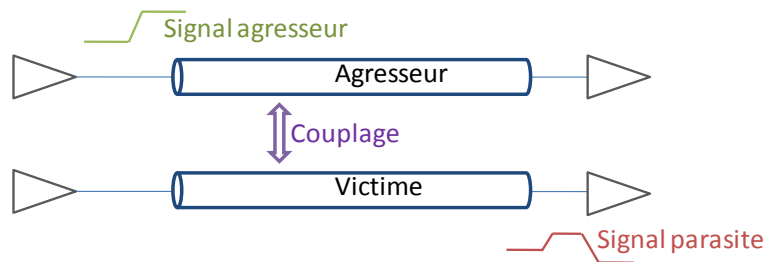


Figure 1-4 : Erreur logique générée par un phénomène de couplage entre deux interconnexions.

Si l'amplitude de ce signal parasite est suffisamment élevée, elle peut engendrer des dysfonctionnements en termes de niveau logique, provoqués par le déclenchement aléatoire de porte logique et compromettre le fonctionnement du circuit.

## I.2 Les systèmes intégrés

Nous avons vu que les performances des circuits étaient fortement dépendantes des limites technologiques relatives aux interconnexions. Afin de limiter les longueurs de ces interconnexions et répondre à des spécifications de plus en plus difficiles à atteindre, comme l'augmentation des puissances de calcul, le faible coût ou encore la faible consommation d'énergie, des systèmes complets intégrés sur une même puce sont apparus. Le système sur puce ou SOC (System On Chip) répond à la problématique de miniaturisation alors que le système en boîtier ou SIP (System In Package) répond à la problématique de diversification.

### I.2.1 Système sur puce (SOC : System On Chip)

Ce système tout intégré consiste à regrouper des sous-systèmes de types fonctions logiques, analogiques, RF ou encore signaux mixtes sur une même puce [Belleville, 2006]. La conception de l'ensemble des composants se fait à partir du même nœud technologique, ce qui implique des procédés de fabrication compatibles et l'obligation d'utiliser la technologie du sous-système le plus évolué (d'un point de vue technologique) pour l'ensemble de la puce.

Les atouts d'un tel système sont : la performance, la fiabilité, la compétitivité et la difficulté à le copier. Ses limitations sont, dans un premier temps, financières. A partir du nœud 65 nm il est inutile de concevoir un SOC à moins de quelques dizaines de millions d'unités par an. De plus, intégrer un maximum de fonctions sur un même design devient extrêmement complexe du point de vue du savoir faire et des moyens de calcul. Enfin, plus le système regroupe de fonctionnalité, plus sa taille augmente, ce qui implique des interconnexions plus grandes et donc des performances amoindries.

### I.2.2 Système en boîtier (SIP : System In Package)

Le SIP associe dans un même boîtier des éléments très hétérogènes : circuits intégrés, MEMS, batteries, composants RF, modules de traitement biologique... Ces éléments seraient impossibles à fabriquer ensemble car les procédés sont trop hétérogènes. Il s'agit donc d'assembler ces différentes puces, puis de réaliser la connectique pour les relier. La conception est simplifiée, les masques sont beaucoup moins chers, mais le coût de production unitaire est élevé et l'ajout *a posteriori* de la connectique pénalise la fiabilité et la vitesse des circuits. Les SIP sont la bonne solution quand le délai de mise sur le marché est déterminant ou que les quantités prévues rendent incertaine la rentabilité d'un SOC [Belleville 2006].

Les atouts d'un tel système sont l'intégration de composants de natures réellement différentes et le gain de place sur la carte mère. Dans le cas de connexions par câblage (*wire bonding*) la superposition de plusieurs puces électroniques est possible et dans le cas d'un report (*flip chip*) l'encombrement est moindre et on s'affranchit des pertes de performances engendrées par les câbles de *wire bonding* (Figure 1-5). Les limitations du SIP sont principalement l'encombrement et les pertes de performance dues aux liaisons dans le cas de connexions par *wire bonding* et l'impossibilité de superposer les puces pour le *flip chip*.

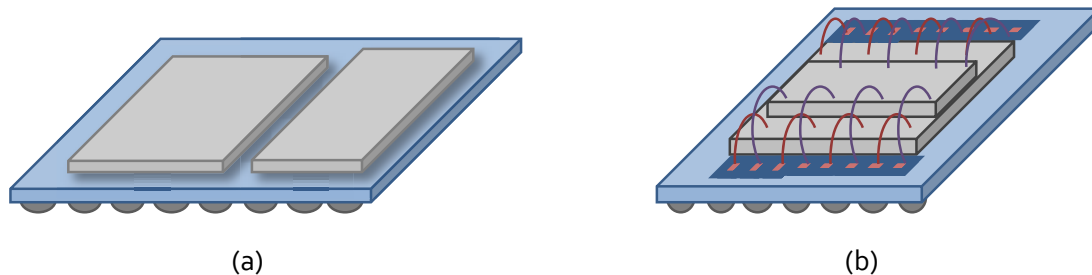


Figure 1-5 : Schémas de connexion. (a) Par report (*flip chip*). (b) Par câblage (*wire bonding*).

Le compromis idéal serait de pouvoir associer à la fois le principe d'encombrement minimal, propre au *flip chip* et le principe de superposition, propre au *wire bonding*. D'où cette tendance depuis quelques années à se tourner vers l'intégration en trois dimensions.

### I.2.3 Vers l'intégration 3D

L'intégration en trois dimensions s'impose tout naturellement comme une alternative présentant de nombreuses solutions aux problèmes causés par une intégration planaire. Elle permet, entre autre, la co-intégration de technologies hétérogènes, l'augmentation de la densité de transistors par unité de surface sans réduire leurs dimensions et de réduire les délais dans les interconnexions en remplaçant de longues interconnexions horizontales par de courtes interconnexions verticales. Elle

réunit toutes les spécificités du *More Moore* par sa tendance à la miniaturisation et du *More than Moore* par sa tendance à la diversification.

La Figure 1-6 illustre parfaitement ces tendances. On remarque que la combinaison des SIP et SOC donne une grande hétérogénéité (visible sur l'axe horizontal) et une forte miniaturisation avec le passage des différentes générations technologiques (visible sur l'axe vertical).

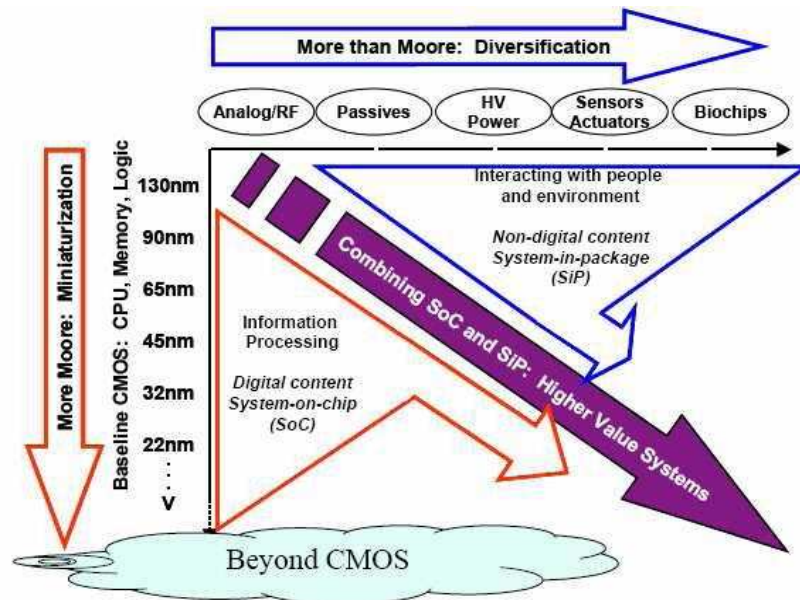


Figure 1-6 : Positionnement de l'intégration 3D dans le contexte de l'industrie des semi-conducteurs [ITRS, 2005].

## II Les options technologiques pour une intégration 3D

L'intégration 3D permet d'augmenter les performances des circuits (comme la bande passante entre un processeur et une mémoire), de réduire la consommation électrique en remplaçant de longues connexions horizontales par de courtes connexions verticales et de baisser les coûts de production en choisissant les technologies adaptées à la fonction globale recherchée.

Tout ceci est possible grâce à un empilement direct de plusieurs couches de circuits (puces, plaques..) reliées électriquement entre elles par des via traversant le silicium, nommées TSV. La constitution physique de ces TSV implique des étapes de fabrication spécifiques, comme la gravure du TSV dans le silicium [Tezcan, 2006] et [Hopkins, 2007], l'isolation des flancs du TSV, qui évite tout risque de court circuit [Chang, 2004], le dépôt du matériau barrière, qui empêche la diffusion du cuivre dans le silicium [Zhang, 2005] et le remplissage du TSV par un matériau conducteur, qui peut être partiel ou complet [Kim, 2006].

Parallèlement le fait de reporter plusieurs puces et de les intégrer suivant un troisième axe, oblige les concepteurs à maîtriser de nombreuses technologies apparues avec l'intégration 3D, comme

l'empilement de différentes puces, l'alignement entre deux puces, le collage ou encore l'amincissement du silicium. Ces notions seront omniprésentes dans la suite des travaux et détaillées dans la partie suivante.

## II.1 Le type d'empilement

La principale différence entre les divers schémas de fabrication d'un circuit 3D est de savoir si l'intégration est effectuée par une approche puce à puce (D2D : die-to-die) [Lo, 2007], puce à substrat (D2W : die-to-wafer) [Pozder, 2008], [Leduc, 2009] ou substrat à substrat (W2W : wafer-to-wafer) [Leduc, 2007] (Figure 1-7).

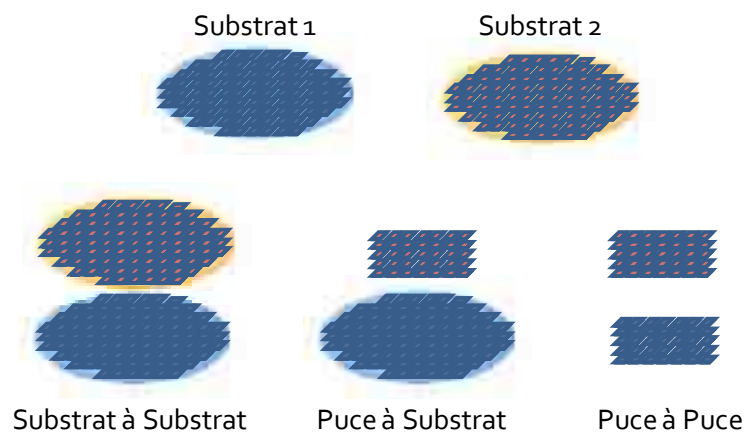


Figure 1-7 : Les trois approches d'empilement 3D.

Le choix de l'approche utilisée va dépendre de plusieurs aspects, comme la spécificité du produit ou la taille des substrats. En effet, si les substrats sont de tailles différentes, seules les approches puce à substrat et puce à puce sont envisageables. De plus si toutes les puces individualisées ont été testées auparavant, le rendement de production par rapport à une technique de report collectif en est considérablement amélioré.

## II.2 L'orientation des circuits

L'orientation du niveau supérieur d'une puce par rapport au niveau supérieur d'une autre puce constitue un point essentiel de l'architecture des circuits 3D. Le schéma d'intégration dans lequel les puces (ou plaques) sont reportées face active contre face active est appelé *Face to Face*, celui dans lequel une face active est reportée contre une face arrière est appelé *Face to Back*, celui dans lequel une face arrière est reportée contre une face active est appelé *Back to Face* et celui dans lequel une face arrière est reportée contre une face arrière est appelé *Back to Back*, (Figure 1-8).



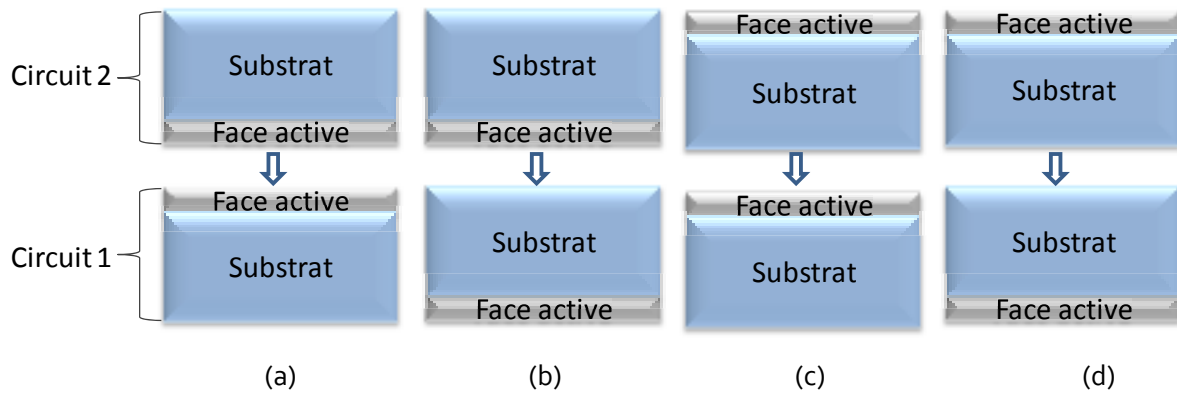


Figure 1-8: Orientation de deux circuits lors d'un empilement 3D.  
 (a) Face to Face. (b) Face to Back. (c) Back to Face. (d) Back to Back

Remarque: Le schéma d'intégration de type Back to Back n'est que très rarement utilisé.

### II.3 L'alignement

La majorité des schémas d'intégration nécessite une grande précision d'alignement. En effet, les TSVs doivent être, après l'étape de collage, en contact avec les plots de contact qui leur correspondent. C'est le diamètre des vias et la taille des plots de contact qui déterminent la tolérance d'alignement. Pour de la moyenne densité d'intégration (diamètre des TSV de 10 à 5  $\mu\text{m}$ ) un désalignement de quelques microns est toléré alors qu'en haute densité d'intégration (diamètre des TSV de 3  $\mu\text{m}$ ) les contraintes d'alignement deviennent inférieures au micron.

### II.4 Le collage

Le collage est généralement réalisé avant l'étape d'amincissement et apporte la stabilité thermomécanique nécessaire à l'interface de collage pour supporter les étapes de fabrication suivantes [Garrou-1, 2008].

Il est possible de regrouper la majeure partie des techniques de collage en deux catégories :

- Avec une couche additionnelle, une couche spéciale est utilisée comme matière intermédiaire pour coller deux niveaux.
- Par collage direct, les diélectriques ou les couches conductrices (principalement des métaux) servent d'interfaces de liaison pour permettre la conduction électrique.

Comme indiqué par la Figure 1-9, deux techniques sont principalement utilisées pour effectuer un collage de type direct (c'est à dire sans couche additionnel) : le collage direct  $\text{SiO}_2/\text{SiO}_2$  ([Stengl, 1989], [Turner, 2002], [Moriceau, 2003], [Di Cioccio, 2005]) et le collage direct par deux surfaces métalliques ([Twordzylo, 1998], [Kim, 2003], [Wang, 2003], [Takahashi, 2003]). Les techniques de collage ayant recours à une couche additionnelle peuvent être regroupées en trois catégories : le

collage par polymère [Kwon, 2006], [Nimura, 2011], le collage dit « eutectique » [Klumpp, 2004] et le collage dit « sol-gel » [Kwon, 2003].

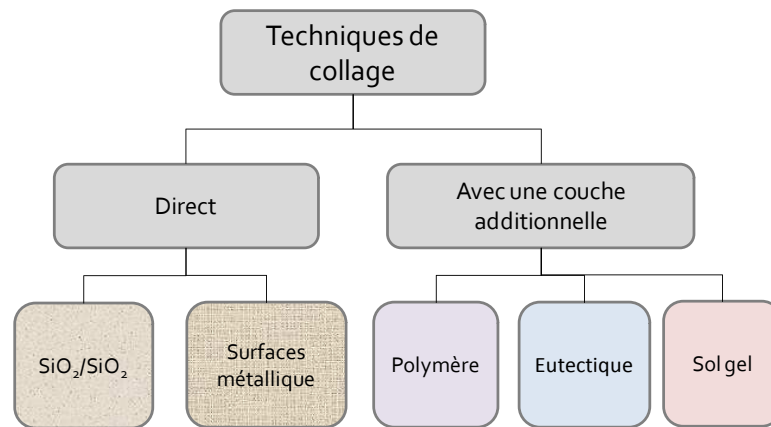


Figure 1-9 : Principes et techniques de collage.

Dans le but d'illustrer les techniques de collage lors du report d'une puce sur une autre, un schéma illustrant les évolutions technologique sur les techniques de collage ainsi que les détails sur le mécanisme de collage direct sont présentés dans les parties suivantes.

#### II.4.1 Evolution des techniques de report de puces.

La Figure 1-10 illustre la forte évolution des techniques de collage en fonction des densités d'intégration. Depuis celles comprenant des métallisations relativement espacées (Billes) à celles consistant à un collage direct de surface de cuivre, en passant par les piliers de cuivre.

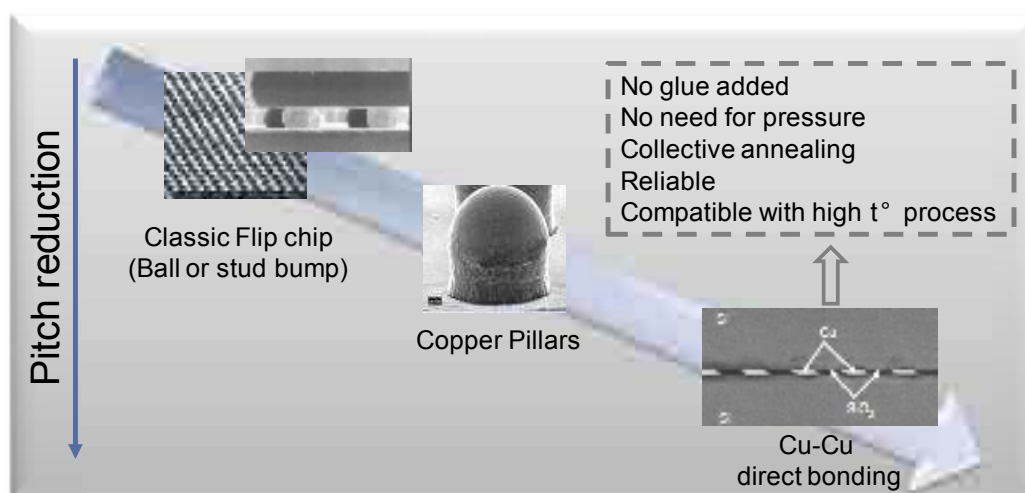


Figure 1-10 : Evolution des techniques de report de puces [Sillon, 2009]

La partie suivante reprend une de ces techniques, la plus intégrée, le collage direct cuivre / cuivre.

## II.4.2 Le collage direct Cuivre / Cuivre

Dans leurs travaux [Taibi, 2010], R. Taibi *et al* présentent des résultats de caractérisation électrique de dispositifs réalisés par un collage direct puce à puce. Le collage a été réalisé à température ambiante et à pression atmosphérique puis il a été suivi d'un recuit de collage de 200 °C ou 400 °C.

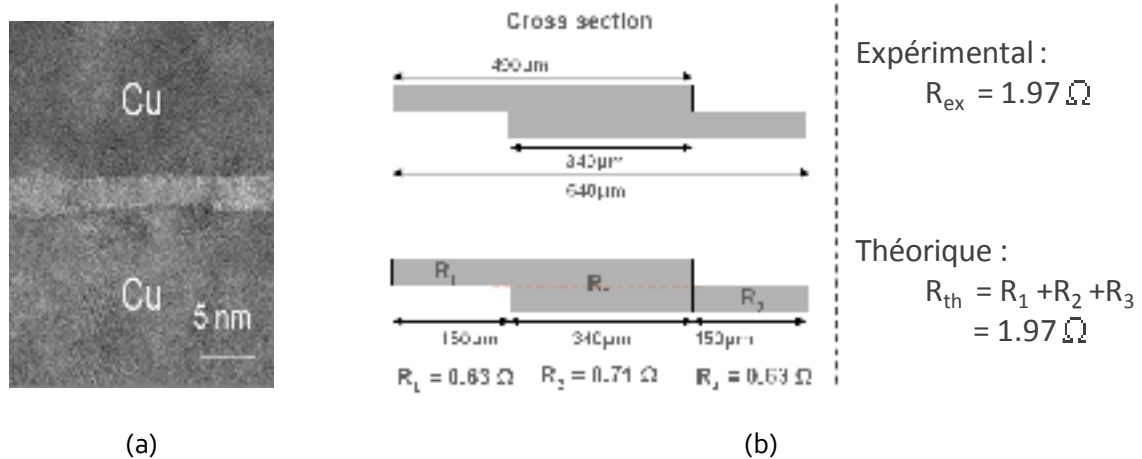


Figure 1-11 : Collage Cu/Cu. (a) Coupe TEM sans recuit.  
(b) Résistances théorique et expérimentale de deux lignes, après un recuit de 400°C.

La Figure 1-11-a est une coupe TEM d'un collage direct Cu/Cu, sans recuit, sur laquelle on observe une interface d'oxyde de cuivre de 4 nm. Après un recuit à 400°C cette interface n'a pas d'influence sur le comportement électrique. La Figure 1-11-b indique que la valeur résistive expérimentale est identique à la valeur résistive théorique d'un dispositif sans interface. Le collage direct Cuivre/Cuivre fait donc partie des clefs technologiques nécessaire à un procédé d'intégration haute densité.

## II.5 L'amincissement du silicium

Les substrats ou les puces collées doivent être amincies, généralement en dessous de 100 μm de manière à pouvoir réaliser les TSV par la suite. L'étape d'amincissement est très difficile à réaliser d'autant plus avec l'augmentation des diamètres des plaques. Elle consiste, dans un premier temps, à arracher progressivement le surplus d'épaisseur du substrat par action mécanique puis d'effectuer un amincissement de manière plus fine par action mécano-chimique. Actuellement, la plupart des procédés 3D IC visent des épaisseurs de substrat considérablement plus minces que 100 μm. C'est pour cette raison que des poignées temporaires sont souvent utilisées pour réaliser les étapes d'amincissement [Garrou-2, 2008].

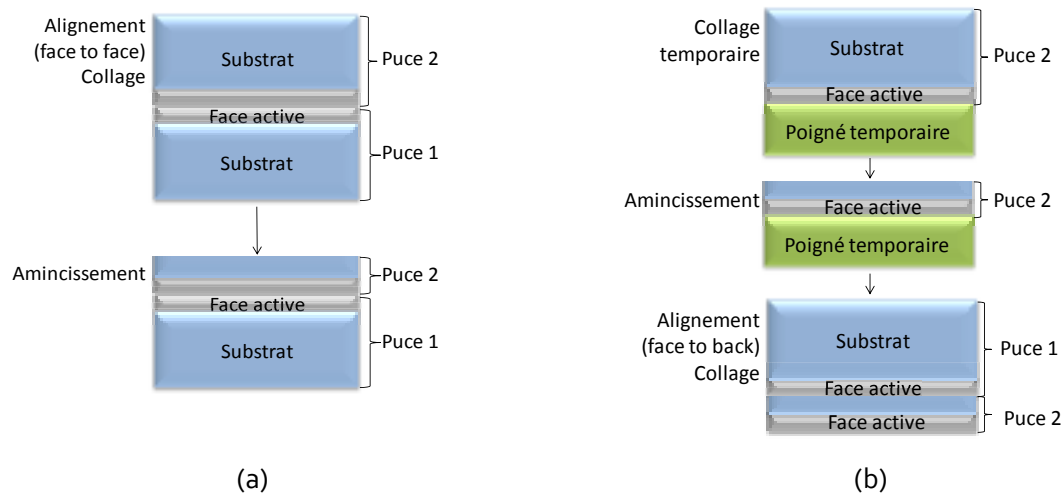


Figure 1-12 : Amincissement du substrat lors d'un collage plaque à plaque.  
 (a) Direct en Face to Face (b) Avec une poignée temporaire en Face to Back.

La Figure 1-12 schématise les étapes menant à l'aminçissement d'un substrat de silicium dans le cas d'un collage plaque à plaque. Pour une orientation Face to Back, le maintien mécanique des plaques durant l'aminçissement du silicium et les étapes d'intégration en face arrière qui suivent sont assurés par la présence d'une poignée temporaire, généralement constituée d'une plaque de silicium ou de verre maintenue au moyen d'une colle.

Dans le cas d'un collage puce à puce ou de puce à plaque il faut également une poignée temporaire quelque soit l'orientation.

### III Les interconnexions en intégration 3D

La transmission de signaux au travers d'un empilement multi-puces nécessite l'étude de chaque brique élémentaire d'interconnexion servant à la transmission d'un signal dans un environnement 3D.

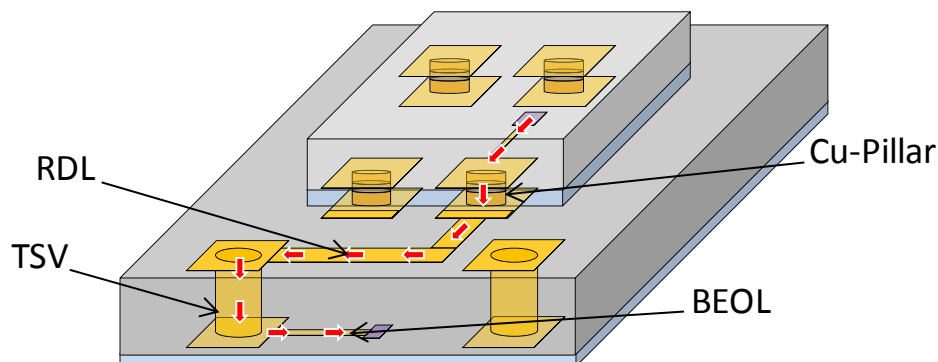


Figure 1-13 : Exemple de chemin parcouru par un signal lors de l'empilement de deux puces.

Dans l'exemple présenté Figure 1-13, la chaîne globale d'interconnexion est composée de plusieurs briques élémentaires d'interconnexions, spécifiques à une intégration 3D, de topologies et de caractéristiques différentes. Ces briques, horizontales lorsqu'elles appartiennent au BEOL et à la couche de redistribution (RDL) et verticales lorsqu'elles sont de type via traversant le silicium (TSV) ou de type pilier de cuivre (Cu-pillar), doivent être, dans un premier temps, étudiées individuellement pour ensuite permettre l'étude de la chaîne entière.

### III.1 Les interconnexions

Sont nommées briques élémentaires, différents types d'interconnexions permettant la transmission d'un signal dans un environnement 3D. La Figure 1-14 met en évidence ces briques élémentaires nécessaires à la transmission d'un signal entre deux puces et entre une puce et un substrat. Les dimensions des dispositifs indiquées appartiennent à la génération dite de moyenne densité d'intégration.

Que ce soit pour une communication puce à puce ou pour une communication puce à substrat, les interconnexions ou briques élémentaires nécessaires à la transmission d'un signal sont composées d'interconnexions du BEOL, d'interconnexions de RDL, de TSV et de Cu-Pillar. Dans cette partie, ces interconnexions ainsi que des fonctionnalités qui leurs sont associées (inductances planaire, TSV à structure coaxial ou encore capacité TSV) sont présentées. Pour finir, certaines applications visées, de type communication entre une mémoire et un processeur ou communication entre un processeur et un substrat sont exposées.

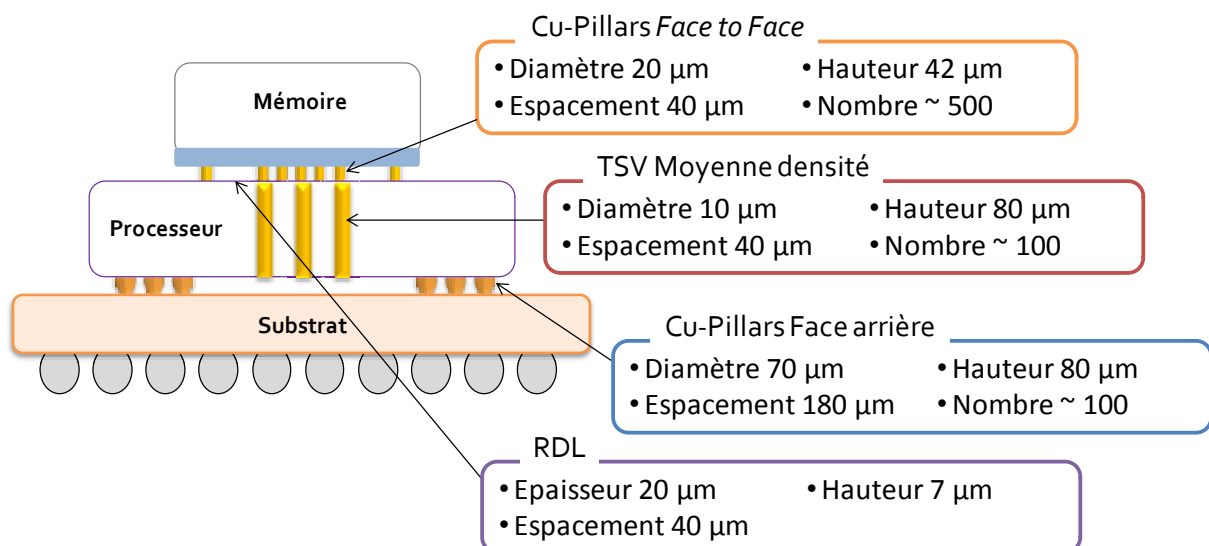


Figure 1-14 : Briques élémentaires d'interconnexions dans des liaisons puce à puce et puce à substrat.

### III.1.1 Les interconnexions de la couche de redistribution (RDL)

Dans la littérature, on rencontre des interconnexions de RDL élaborées suivant deux procédés (Figure 1-15). Le premier est généralement utilisé pour des technologies à faible ou moyenne densité, les interconnexions sont intégrées dans une résine épaisse par un procédé de dépôt électrochimique ECD (Electro Chemical Deposition). Le second procédé est principalement associé à des technologies de haute densité d'intégration, les interconnexions sont intégrées dans une cavité damascène par un procédé ECD suivie d'une « planarisation » mécano chimique, CMP (Chemical Mechanical Polishing).

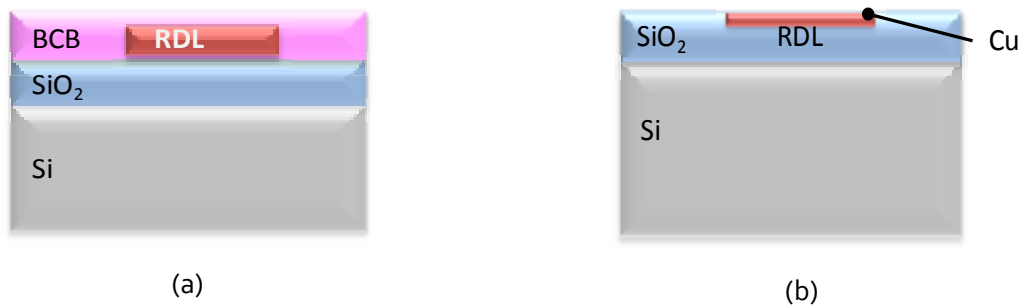


Figure 1-15 : Schémas des empilements comprenant des interconnexions de RDL.  
(a) Dans une résine épaisse. (b) Dans une cavité damascène.

### III.1.2 Les interconnexions de la couche de Back End Of Line (BEOL)

Le BEOL ainsi que son réseau d'interconnexions ont déjà été largement présentés dans la première partie de ce chapitre, néanmoins la vue schématique d'une interconnexion comprise dans le BEOL d'une puce est représentée sur la Figure 1-16.

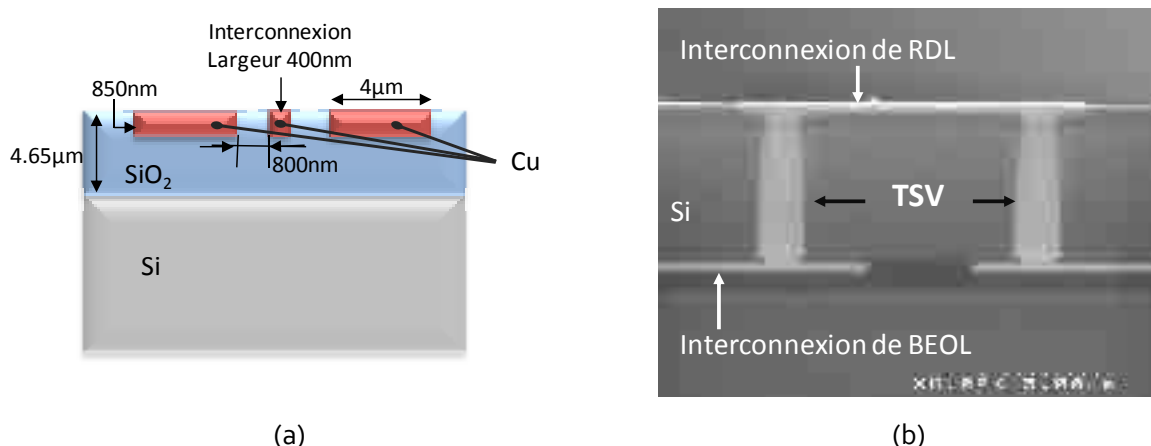


Figure 1-16 : Interconnexion comprise dans le BEOL. (a) Vue schématique (b) Coupe TEM.

Les dimensions annoncées sont celles d'interconnexions intégrées dans le BEOL, pour une puce, de la génération 45 nm.

### III.1.3 Les via traversant le silicium (TSV)

Les TSV peuvent être classifiés suivant quatre flots d'intégration, (Figure 1-17). Le terme « Via First » est utilisé lorsque le TSV est réalisé avant toute autre étape. Si sa réalisation s'effectue entre les étapes du FEOL (Front End Of Line) et du BEOL le terme « Via Middle » est employé. Enfin, lors d'une réalisation effectuée après l'étape du BEOL ou après l'étape de collage des puces, on parlera respectivement de « Via Last » et de « Via After Bonding ». Le choix des flots d'intégration des vias se fait en fonction de l'application visée pour le circuit.

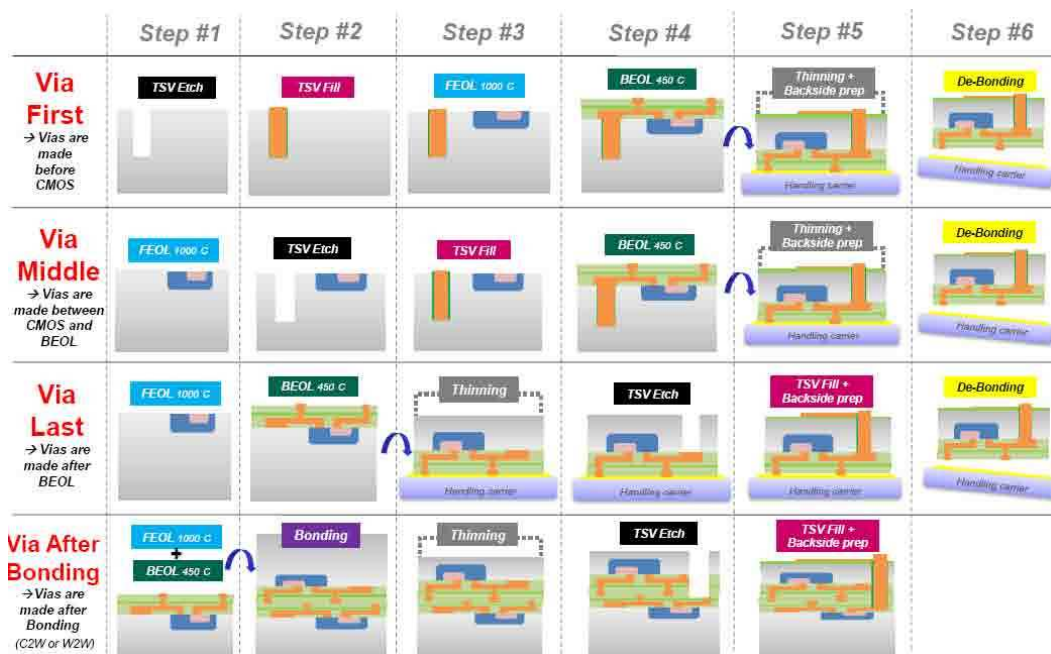


Figure 1-17 : Illustration des principaux scénarios d'intégration 3D. [Yole, 2009]

Dans mes travaux, aucun TSV issu d'un flot d'intégration de type Via First n'est étudié. Ce choix a été motivé par deux raisons. La première étant le flot d'intégration lui même, très peu flexible et dont la liste des matériaux utilisables pour le remplissage est restreinte (la contamination durant la fabrication des transistors banni le cuivre). La seconde, vient de l'impossibilité de les intégrer dans une grande majorité d'application. La Figure 1-18, illustre ces propos en montrant les TSV possibles à intégrer, ceux fréquemment observés et ceux qui le seront avec certitude dans le futur, en fonction des applications visées.

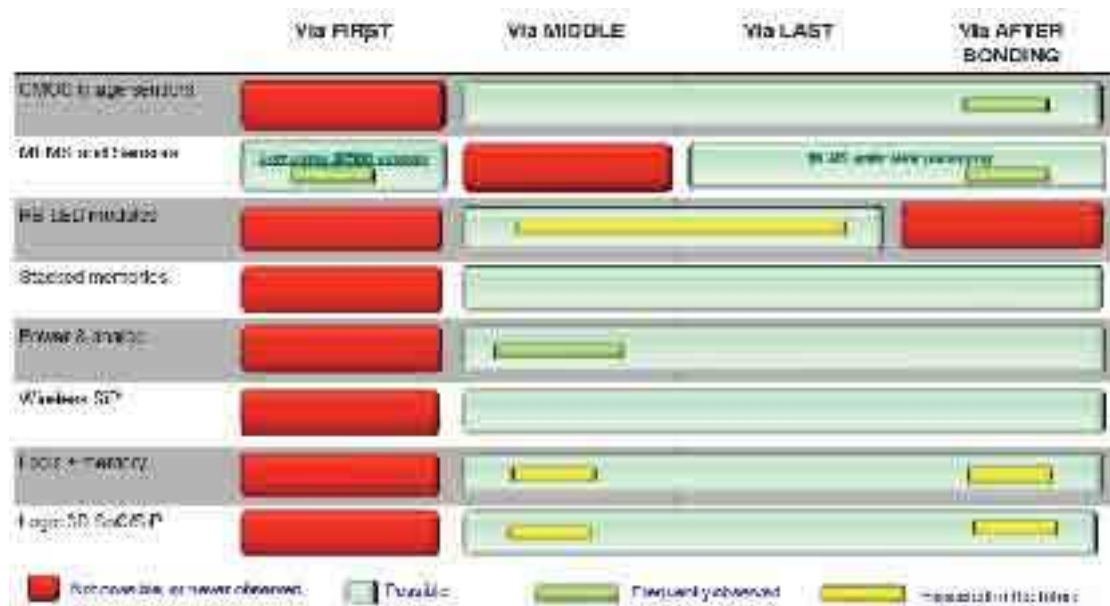
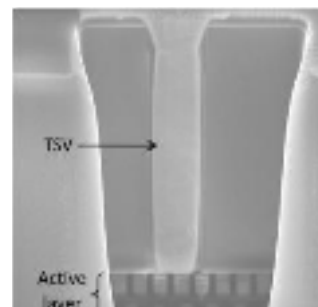


Figure 1-18 : Types de vias recommandés en fonction des applications visées. [Yole, 2009]

Parmi les TSV étudiés deux principales technologies sont utilisées. Dans la première, le TSV est fabriqué après le collage [Druais, 2010] et n'est que partiellement rempli (Figure 3-19-a). Dans la seconde technologie les TSV sont réalisés avant le collage, ils sont nommés « TSV Middle » et sont complètement remplis de cuivre (Figure 3-19-b).



(a)



(b)

Figure 3-19 : Coupe SEM de TSV (a) Partiellement remplie. (b) Complètement remplie.

### III.1.4 Les piliers de cuivre (Copper-Pillar)

Les Cu-Pillars sont classifiés, en dimension et composition, en fonction de la place qu'ils occupent. En effet, comme indiqué sur la Figure 1-20, ils peuvent servir soit à relier deux puces entre elles ( $\mu$ Bump), soit à relier une puce avec un substrat (Bump).



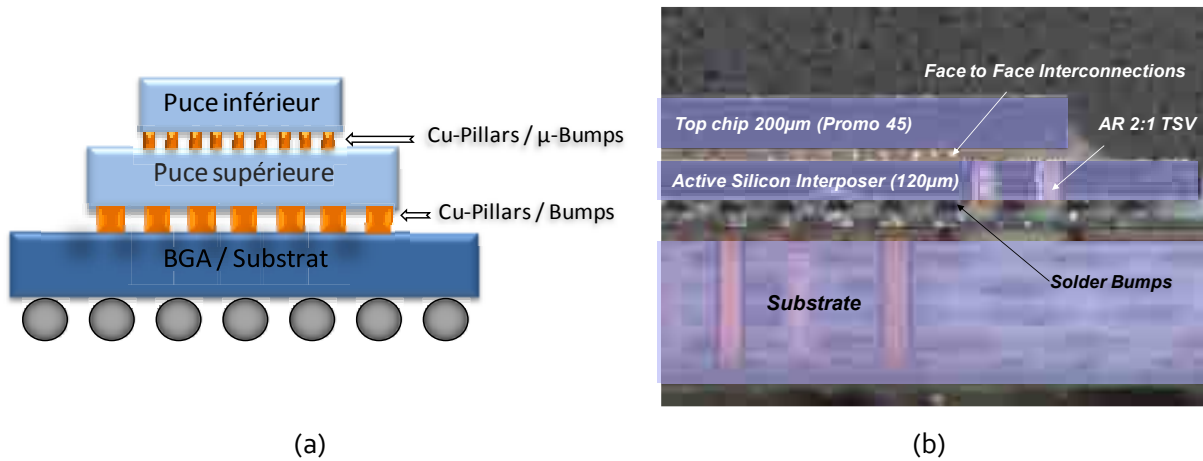


Figure 1-20 : Position des Cu-Pillar lors d'un empilement multi-niveaux.  
(a) Vue Schématique. (b) Photo MEB [Cheramy, 2009].

Les  $\mu$ Bumps sont constitués de trois couches, deux en cuivre séparées par une couche de SnAg alors que le Bump n'est constitué que d'une couche de cuivre et une couche de SnAg.



Figure 1-21 : Coupe MEB de Cu-Pillar. (a) Utilisé lors d'un collage puce à puce.  
(b) Utilisé lors d'un collage puce à substrat.

### III.2 Fonctions associées à ces interconnexions.

La diversité de ces interconnexions tant au niveau de leurs géométries qu'aux niveaux de leurs générations favorise l'apparition de fonctions innovantes et variées. Dans cette partie sont présentées quelques unes des fonctionnalités réalisables grâce aux interconnexions énoncées précédemment.

Dans un premier temps nous présentons une inductance planaire située dans le BEOL. Cette fonctionnalité n'est pas nouvelle, mais il est intéressant de déterminer l'évolution de ses performances lorsque qu'elle est plongée dans un environnement 3D. Puis nous verrons une inductance située au niveau de la couche de redistribution, qu'elle soit simple ou doublée avec une inductance de BEOL, faisant elle, partie des nouvelles fonctionnalités rendues possible par l'intégration 3D. Pour finir nous présenterons un autre exemple de fonctionnalité propre à l'intégration 3D, des TSV de structure coaxiale.

### III.2.1 Inductance au niveau BEOL

L'inductance présentée Figure 1-22 sera utilisée pour mettre en évidence l'impact de l'empilement 3D sur les performances des inductances lors de l'empilement de deux puces en architecture Back to Face et en architecture Face to Face.

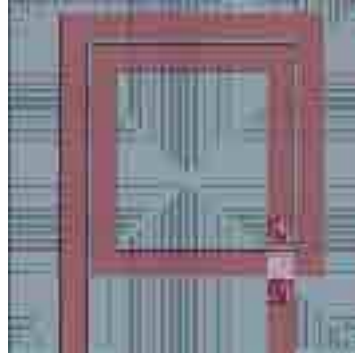


Figure 1-22 : Schéma d'inductance intégrée dans le BEOL.

Ces inductances sont largement intégrées pour des applications d'amplificateur, de filtre, d'oscillateur etc... Elles sont de type planaire et sont intégrées sur les niveaux d'interconnexions supérieurs.

### III.2.2 Inductances aux niveaux RDL.

L'inductance présentée Figure 1-23-a est intégrée sur le niveau RDL. L'inductance présentée Figure 1-23-b est double et elle est constituée de deux inductances similaires, une sur le niveau RDL et une sur le niveau BEOL, relié par deux TSV.

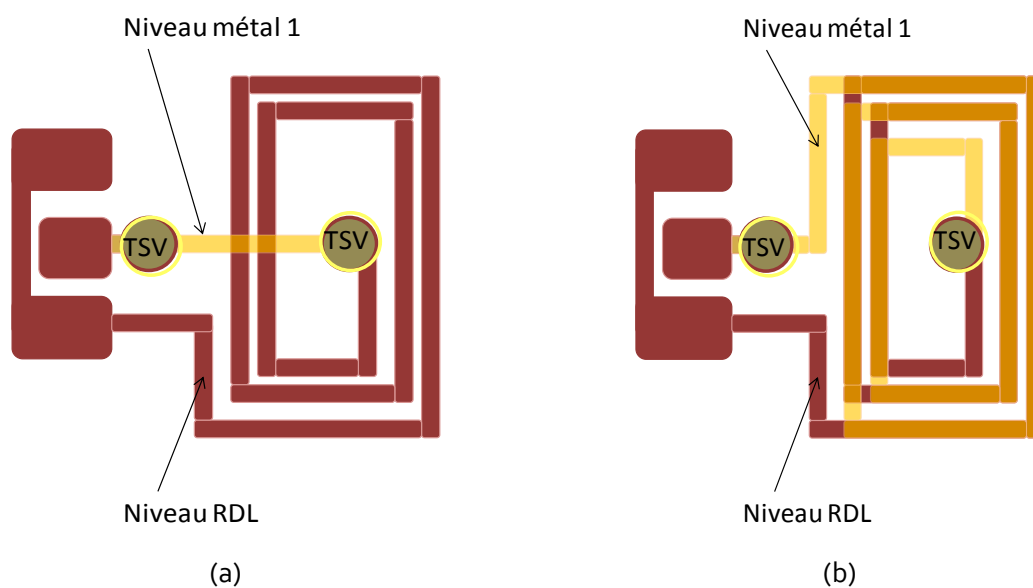


Figure 1-23 : Schémas d'inductances intégrées sur la couche RDL.  
(a) Inductance simple. (b) Inductance double.

Des mesures en hyperfréquence nous permettrons dans le chapitre 2 de comparer le facteur de qualité d'une inductance simple à celui d'une inductance double.

### III.2.3 TSV en structure coaxiale [Ho, 2008]

Un TSV coaxial est réalisé à l'intérieur d'un substrat de silicium. Une barrière d'isolation est déposée sur le flanc du substrat. Une couche de métal conducteur représentant la masse est formée entre cette barrière et une épaisse couche de diélectrique. Au centre du diélectrique un noyau conducteur représente le signal (Figure 1-24).

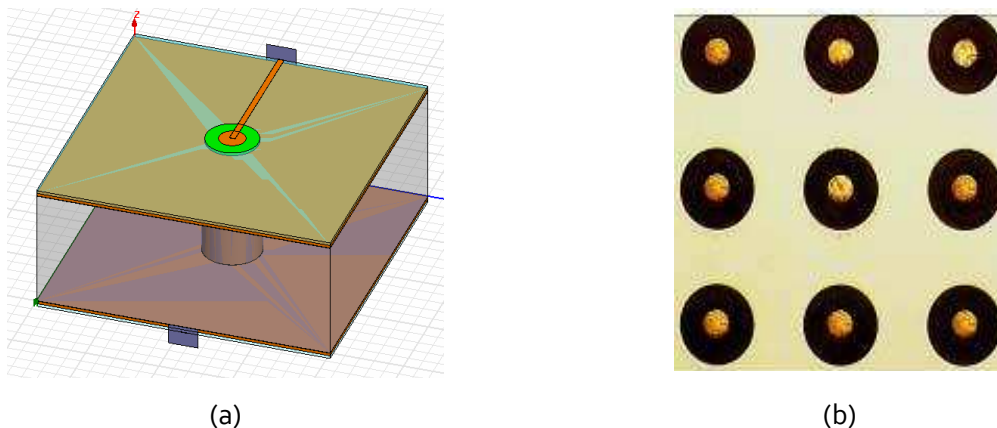


Figure 1-24 : TSV coaxial. (a) Schéma (b) Image optique.

Les premiers résultats sur cette fonctionnalité ont montré que les TSV coaxiaux sont capables de minimiser les pertes dans le substrat (fréquence de 10 GHz). Ces structures, très large bande, présentent néanmoins un inconvénient majeur au niveau des connexions. Il est en effet très difficile d'amener le signal au centre du TSV sans toucher la masse se trouvant sur ses flancs.

### III.2.4 Triple condensateur MIM par TSV [Roozeboom, 2008].

Les futures générations de cellules émetteurs-récepteurs RF nécessiteront un haut niveau d'intégration, de préférence en utilisant la troisième dimension. Les System-in-Package (SiP) présentés par Fred Roozeboom ont montré des structures de condensateurs dans des TSV présentant un record en termes de densité de capacité et de tension de claquage.

Le principe est celui d'une triple capacité MIM. La Figure 1-25-a montre l'empilement des différentes couches en planaire. La première couche est composée de dioxyde de silicium de 5 nm. Sur cette couche sont déposées successivement des couches de TIN (trois au total de 20 nm chacune) et d' $\text{Al}_2\text{O}_3$  (deux couche de 10 nm chacune). Ce principe est, grâce à un masque dédié, appliqué à une triple capacité MIM par TSV (Figure 1-25-b). Les électrodes C1, C2 et C3 permettant le

contact électrique avec l'extérieur. La très bonne adhésion des différentes couches tout au long du Via est représentée Figure 1-25-c.

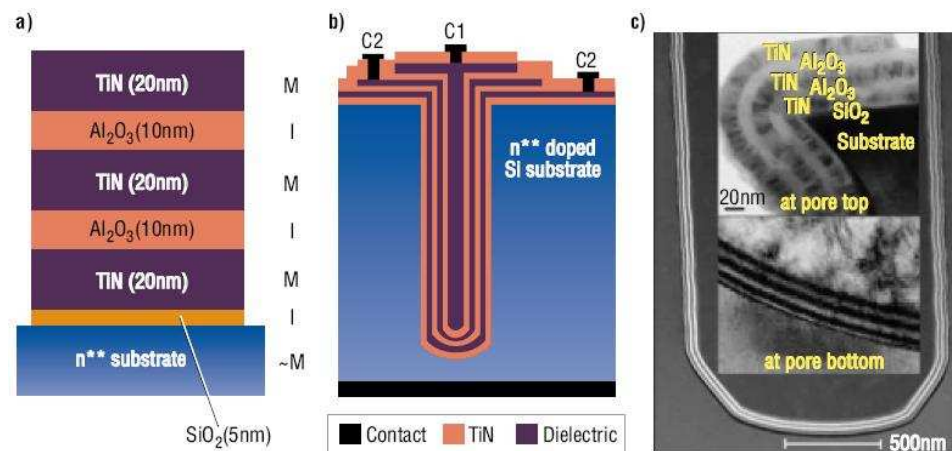


Figure 1-25 : Schéma d'un triple condensateur MIM  
 a) composé de trois couches de TiN et de trois couches de diélectriques,  
 b) test d'un condensateur par TSV (les contacts des électrodes TiN C1-C3 sont à l'avant),  
 c) Coupe TEM de l'empilement des couches du condensateur MIM.

Les mesures de capacité réalisées à 10kHz, ont démontré un record en termes de densité de capacité 400nF/mm<sup>2</sup> et de tension de claquage supérieure à 6V. La moyenne a été effectuée sur plus de 35 condensateurs par TSV et avec une surface d'électrode de 100 × 100 μm<sup>2</sup>.

### III.3 Applications envisagées

Nous avons vu précédemment que la miniaturisation et l'hétérogénéité des systèmes atteignent leurs limites de réalisation, lors d'une intégration traditionnelle 2D. Cependant les spécificités demandées et les vitesses de communication entre les mémoires et les processeurs ou entre les processeurs et les substrats ne cessent d'augmenter.

La Figure 1-26, montre que l'intégration 3D offre dans ce cas de véritables solutions, tant au niveau de la consommation, que de l'encombrement ou encore de la bande passante.

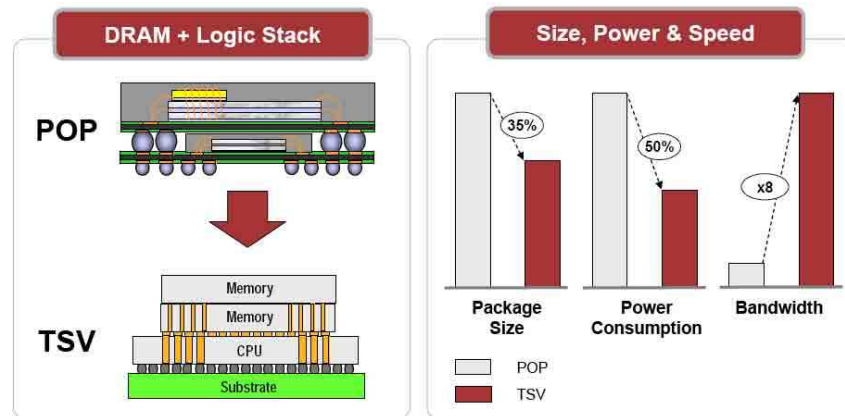


Figure 1-26 : 3D packaging vs 3D intégration.

Une intégration 3D avec TSV, en comparaison avec une architecture « package on package » POP, où les puces sont préalablement mises en boîtier avant d'être connectées, diminue l'encombrement de 35%, la consommation d'énergie de 50% et permet de multiplier jusqu'à huit la bande passante.

La Figure 1-27 schématise un exemple d'application intégrant trois puces. Dans le cas présent il s'agit d'une mémoire Wide I/O reportée Face to Back sur un processeur « System On Chip », lui même reporté en Face to Face sur le BGA (Ball Grid Array) noté « Package Substrate ». Cet empilement est pressenti aujourd'hui comme l'application principale de l'intégration 3D pour les toutes prochaines années.

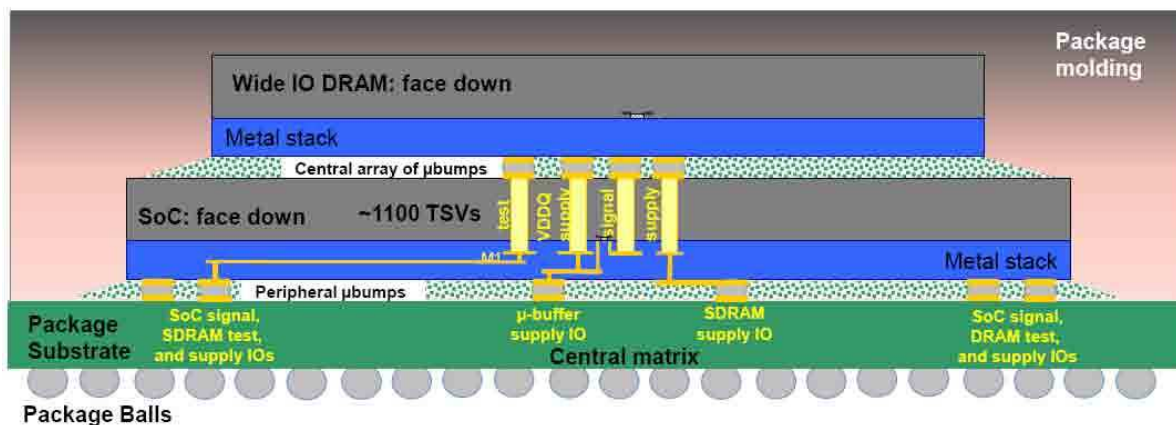


Figure 1-27 : Technologie Wide I/O [Kimmich, 2011]

Les spécifications d'une telle application sont connues, la mémoire Wide I/O communique avec le processeur à une fréquence maximale de 200 MHz. Le processeur (SOC) communique, quand à lui avec le BGA à une fréquence maximale de 5.8 GHz.

---

## IV Mes travaux de thèse

L'objectif premier de ce travail de thèse est l'analyse et l'optimisation des performances électriques des réseaux d'interconnexion et des composants passifs lors d'un empilement 3D.

Pour se faire des briques élémentaires, constituées d'interconnexion et de composant passif ont été caractérisées et modélisées dans les chapitres 2 et 3. Des séries d'études dans les domaines fréquentiel et temporel ont été menées afin de prédire le comportement et les performances de ces différents éléments.

Le chapitre 4 met en scène plusieurs scénarios afin de déterminer le meilleur chemin pour router les signaux, le meilleur empilement suivant l'application visée et cibler l'élément le plus critique dans une chaîne d'interconnexion en fonction des densités d'intégration. Ceci afin de dégager de grands axes permettant un design optimisé des systèmes électroniques intégrés en trois dimensions (3D).

Pour finir, l'application principale de l'intégration 3D, le Wide I/O a été étudiée dans le chapitre 5, en fonction de diverses options technologiques, dans le but de déterminer en quelles mesures les spécifications attendues en termes de performances électriques pour une telle application peuvent être atteintes.

# Chapitre 2

## Caractérisation et modélisation des briques élémentaires d'interconnexions dans les empilements 3D de circuits intégrés

---

Ce chapitre est dédié à la caractérisation et à la modélisation électrique de dispositifs passifs intégrés, telles que les interconnexions, dans le domaine des hautes fréquences (typiquement des MHz à quelques dizaines de GHz).

La première partie expose les spécificités des mesures hyperfréquences sur des composants passifs enfouis au sein d'un circuit intégré. En effet toute tentative de caractérisation de briques élémentaires d'interconnexions de circuit intégré nécessite un protocole expérimental bien particulier. Les faibles dimensions des dispositifs testés et la grande précision des mesures nécessaires à une bonne extraction d'un modèle électrique équivalent de ces dispositifs, conduisent à développer un outil indispensable à la caractérisation, la procédure dite d'épluchage ou « de-embedding ». Elles permettent d'extraire les caractéristiques électriques intrinsèques d'un composant enfoui dans un environnement complexe. Plusieurs procédures de de-embedding utilisées dans ces travaux sont présentées, chacune ayant ses avantages et inconvénients.

Dans la seconde partie les briques élémentaires d'interconnexions présentées dans le chapitre 1 sont caractérisées et modélisées à l'aide des méthodes de de-embedding citées ci-dessus. Les résultats expérimentaux obtenus par ces différentes méthodes d'extraction sont confrontés à des simulations, effectuées avec des codes de calculs électromagnétiques tels que HFSS et Maxwell 2D d'Ansoft, dans le but de les valider et les utiliser pour des études paramétriques dans la suite de nos travaux.

## Sommaire

<b>I</b>	<b>INTRODUCTION.....</b>	<b>36</b>
<b>II</b>	<b>OUTIL INDISPENSABLE A LA CARACTERISATION : LE DE-EMBEDDING .....</b>	<b>37</b>
II.1	METHODE DE DE-EMBEDDING GENERIQUE .....	39
II.1.1	Rappel sur la méthode développée par E.P. Vandamme.....	39
II.1.2	Méthode de E.P. Vandamme dérivée et adaptée à nos contraintes .....	40
II.1.2.a	Etape de modélisation.....	40
II.1.2.b	Etape de détermination des éléments parasites $G_i$ , $Z_i$ , $Y_i$ .....	41
II.1.2.c	Etape d'épluchage ou de-embedding du DST .....	43
II.1.2.d	Conclusion .....	44
II.2	METHODE DE DE-EMBEDDING DESTINEES A LA CARACTERISATION DE L'EXPOSANT DE PROPAGATION D'UNE LIGNE DE TRANSMISSION.....	44
II.2.1	Modèle des lignes de transmission .....	44
II.2.2	Extraction de l'exposant de propagation $\gamma$ .....	46
II.3	METHODE DE DE-EMBEDDING DESTINEES A LA CARACTERISATION DE L'IMPEDANCE CARACTERISTIQUE D'UNE LIGNE DE TRANSMISSION.....	47
II.3.1	Méthode basée sur les travaux de D.F. Williams .....	48
II.3.1.a	Etape de modélisation.....	49
II.3.1.b	Etape de détermination des termes d'erreur .....	49
II.3.1.c	Conclusion .....	51
II.3.2	Méthode DUMMC (De-embedding using modification of materials characteristics). .....	52
II.3.2.a	Etape de modélisation.....	52
II.3.2.b	Etape d'épluchage ou de-embedding du DST .....	56
II.3.2.c	Conclusion .....	57
II.3.3	Méthode d'épluchage par bloc .....	57
II.3.3.a	Présentation générale .....	57
II.3.3.b	Etape de mise en équations et de modélisation .....	59
II.3.3.c	Etapes de mesures .....	60
II.3.3.d	Conclusion .....	62
II.3.4	Méthode de de-embedding de TM. Winkel améliorée .....	62
II.3.4.a	Présentation générale .....	62
II.3.4.b	Mise en équations .....	63
II.3.4.c	Domaine de validité .....	67
II.3.4.d	Conclusion .....	68
II.3.5	Conclusion générale sur les méthodes de de-embedding présentées .....	68



<b>III</b>	<b>CARACTERISATION DES BRIQUES ELEMENTAIRES D'INTERCONNEXIONS DANS UN EMPILEMENT 3D DE PUCES .....</b>	<b>69</b>
III.1	INTERCONNEXIONS DE TYPE RDL (REDISTRIBUTION LAYER). ....	69
III.1.1	Description du dispositif de test associé à une interconnexion de type RDL .....	69
III.1.2	Résultats d'extraction et de modélisation .....	70
III.1.3	Interconnexion de RDL dans un environnement réel d'intégration 3D .....	71
III.1.3.a	Description du dispositif sous test et de sa cellule de test.....	71
III.1.3.b	Résultats.....	71
III.2	INTERCONNEXION DU BEOL ENFOUIE DANS UN ENVIRONNEMENT 3D .....	74
III.2.1	Description du dispositif sous test .....	74
III.2.2	Résultats d'extraction et de modélisation.....	75
III.3	THROUGH SILICON VIA (TSV) HAUTE DENSITÉ.....	77
III.3.1	Description de la cellule de test mesurée .....	77
III.3.2	Méthode d'épluchage appliquée pour la caractérisation du TSV .....	78
III.3.3	Extractions dans les domaines fréquentiel.....	79
III.4	CU-PILLAR.....	81
III.5	INDUCTANCES PLANAIRES INTEGREES.....	81
III.5.1	Description des inductances intégrées sous test.....	82
III.5.2	Résultats d'extraction .....	82
<b>IV</b>	<b>CONCLUSION .....</b>	<b>84</b>

## I Introduction

La quantification, la prédiction et l'optimisation des performances des chaînes complètes d'interconnexions dans un environnement 3D, ne sont possibles que si chaque brique élémentaire d'interconnexion, constituant cette chaîne, a été caractérisée et modélisée individuellement au préalable. Ce chapitre est donc consacré à l'étude de chaque brique élémentaire d'interconnexion pouvant être présente dans une chaîne d'interconnexion, ces briques sont le via traversant le silicium, le pilier de cuivre, la RDL... Un autre composant passif élémentaire, l'inductance planaire, est également étudié pour observer et optimiser ses performances dans un environnement 3D.

Dans une première partie nous avons choisi d'exposer la problématique de la caractérisation électrique, définie au sens de la mesure, des différentes briques élémentaires d'interconnexions répertoriées dans le chapitre 1. En effet ces éléments sont toujours enfouis au sein d'un circuit, que ce dernier soit un circuit intégré opérationnel ou un circuit servant de véhicule de test. Ce caractère enfoui exprime le fait que l'accès électrique à cet élément est complexe, constitué de plusieurs profils de connexions. Par exemple pour accéder à un TSV depuis les bornes de connexion d'un instrument de mesure (analyseur de réseau, VNA), le signal de test va devoir successivement se propager sur :

- un câble coaxial
- une sonde de mesure (probe)
- un plot de contact permettant de connecter la sonde au circuit
- des lignes de transmission intégrées amenant le signal du plot de contact jusqu'au TSV

La Figure 2-1 illustre cet exemple et ce caractère enfoui d'un composant sous test dans un circuit.

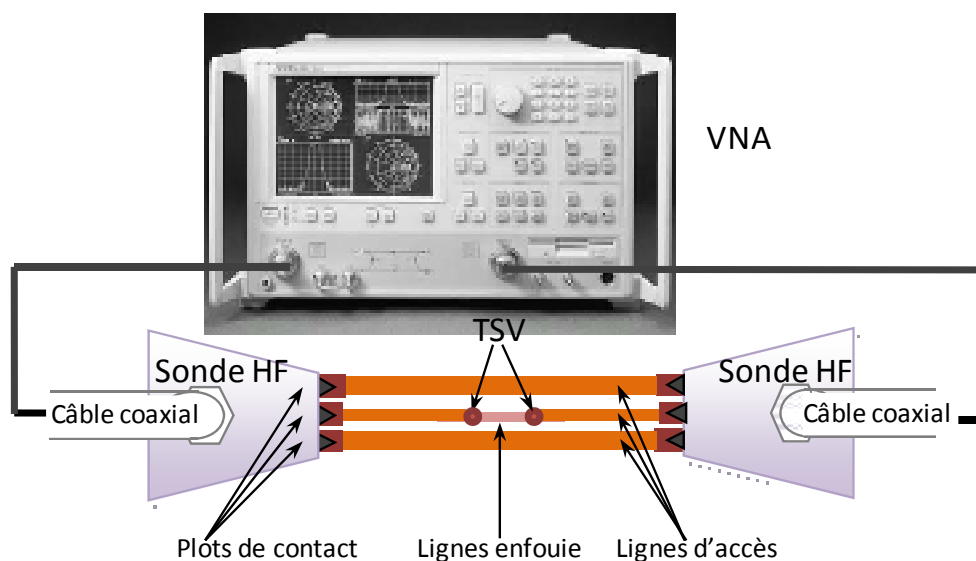


Figure 2-1 : Analyseur de réseaux, raccordé au dispositif mesuré.

L'opération de de-embedding consiste à extraire les caractéristiques intrinsèques du dispositif sous test (TSV dans l'exemple) à partir de la mesure globale réalisée aux bornes de la cellule de test. On s'affranchit ainsi des perturbations créées par l'environnement de test et par les accès au dispositif sous test lui-même. Cette opération est primordiale pour caractériser un composant intégré et d'extraire un modèle électrique équivalent et précis pour ce dernier.

La littérature scientifique est très riche de techniques de de-embedding, souvent associées aux techniques de calibrage des instruments de mesure. Parmi ces techniques, rares sont celles qui sont adaptées aux mesures sur les circuits intégrés et elles sont encore plus rares quand il s'agit d'extraire la matrice chaîne de type ABCD du composant et non uniquement sa matrice de diffusion  $S$ . En effet extraire une matrice chaîne nécessite la connaissance des impédances absolues et non des impédances réduites.

Dans la première partie de ce chapitre nous présentons plusieurs techniques de de-embedding et les travaux menés sur ces dernières. Nous avons amélioré des techniques existantes et développé de nouvelles. Chacune de ces techniques présente des avantages et inconvénients vis à vis des composants que nous avons à caractériser et nous les avons comparés avant de les exploiter. Dans la deuxième partie du chapitre nous donnons les résultats de caractérisation obtenus sur les briques élémentaires d'interconnexion (TSV, RDL,...) et sur des composants passifs intégrés (inductances) et nous les comparons aussi aux résultats de modélisation issus de calculs électromagnétiques.

## II Outil indispensable à la caractérisation : le de-embedding

La phase de de-embedding est similaire à celle nommée calibrage d'un banc de mesure hyperfréquence (Annexe 1). En effet lors de mesures par analyseur de réseau vectoriel ou scalaire une étape de calibrage est nécessaire afin de corriger les erreurs systématiques engendrées par l'instrumentation et l'environnement de mesure. L'opération de de-embedding consiste d'une manière générale à réaliser les opérations suivantes :

- L'adoption d'un modèle électrique qui décrit le comportement de tous les éléments de la cellule de test (CDT) dans laquelle est intégré le dispositif sous test (DST), lui-même représenté par une boîte noire (matrice  $S$ , matrice chaîne, matrice impédance...).
- La détermination des valeurs de tous les éléments de la CDT constituant ce modèle par l'intermédiaire de mesures additionnelles sur des dispositifs complémentaires, dits de De-Embedding et souvent nommés « standard » ou cellule de de-embedding (CDE). Ces derniers doivent témoigner du comportement réel de l'environnement du dispositif sous test.

- L'extraction des paramètres propres ou intrinsèques au dispositif sous test (DST) à partir de la mesure de la réponse globale de la cellule de test (CDT), c'est à dire l'extraction de la réponse du composant seul, une fois soustraient (épluchés) les effets de tout son environnement.

La Figure 2-2 schématise cette notion de de-embedding et ses différentes étapes.

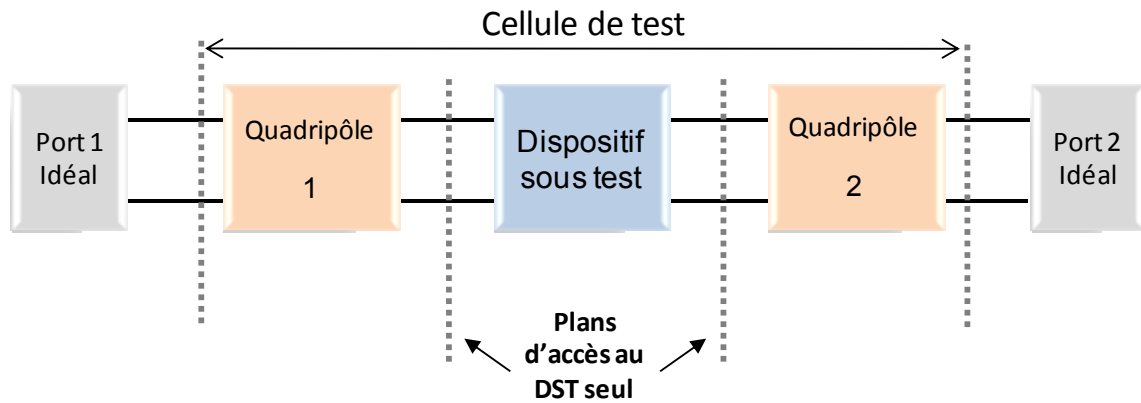


Figure 2-2 : Cellule de test comprenant le dispositif sous test.

De nombreuses méthodes de De-Embedding ont été développées, dont les plus connues et applicables à notre problématique sont [Vandamme, 2001], [Williams, 1995], [Winkel-1, 1996].

On peut dès à présent mentionner les points et définitions suivantes :

- Certaines méthodes de de-embedding s'appliquent à la caractérisation de tous types de composants ou systèmes sous test (par exemple condensateur, inductance, ligne de transmission, transistor, mais aussi filtre, amplificateur...).
- Certaines méthodes de de-embedding s'appliquent uniquement à la caractérisation d'un seul type de composant (par exemple aux lignes de transmission).
- L'étape de de-embedding est souvent menée après une première procédure de calibrage permettant de positionner les plans de mesure au niveau des pointes des sondes de test, c'est à dire aux bornes de la cellule de test (CDT) dans laquelle est enfoui le dispositif à tester (DST). Ce calibrage permet de corriger toutes les erreurs systématiques en amont des sondes de test (VNA, câbles, connecteurs, sondes) et de considérer que la cellule de test (CDT) débute au niveau des plots de contact permettant de relier électriquement le composant à caractériser, que nous nommerons désormais dispositif sous test (DST). La Figure 2-3 décrit cette situation.

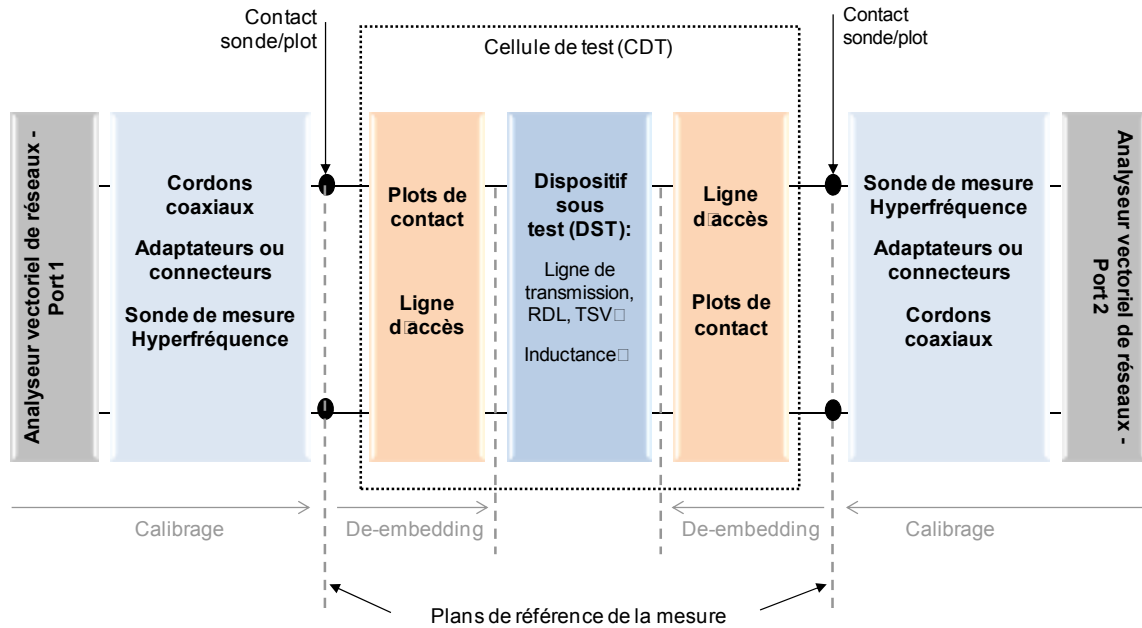


Figure 2-3 : Calibrage et de-embedding d'un dispositif sous test.

## II.1 Méthode de de-embedding générique

Ici une méthode de de-embedding générique, reconnue par la communauté scientifique et très utilisée sur les plateformes de test, utilisable pour tous les types de composant à caractériser (passifs ou actifs), est décrite. Elle a été développée par E.P. Vandamme. Nous proposons ensuite une adaptation de cette méthode à notre problématique de caractérisation de composants passifs intégrés.

### II.1.1 Rappel sur la méthode développée par E.P. Vandamme

Cette méthode repose sur un modèle électrique équivalent rigoureux et générique de l'environnement de test du dispositif à caractériser (DST). Elle autorise des d'accès dissymétriques entre le dispositif sous test et les ports de mesure de la cellule de test, et prend en compte le couplage éventuelle entre ces ports.

La Figure 2-4, représente le modèle électrique équivalent de la cellule de test (CDT) mesurée et dans laquelle est enfoui le dispositif sous test (DST).  $G_1$  et  $Z_1$  correspondent aux effets des accès sur le port 1 du DST,  $G_2$  et  $Z_2$  à ceux sur le port 2,  $G_3$  et  $Z_3$  représentent principalement les couplages entre les deux ports de mesure et du composant sous test.

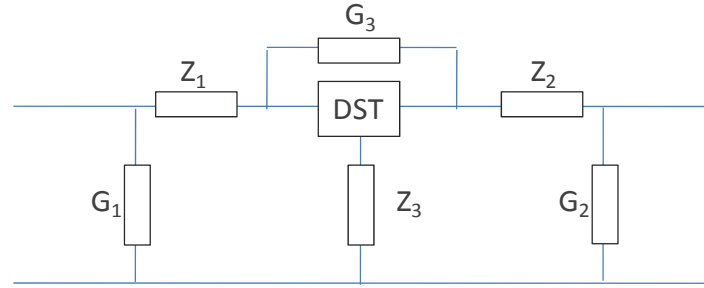


Figure 2-4 : Modèle électrique équivalent de la cellule de test complète (CDT) intégrant le dispositif de test (DST) à caractériser

Le calcul de toutes les impédances « parasites » entourant le DST dans le modèle nécessite la mesure de plusieurs autres dispositifs de test complémentaires, nommés cellule ou standard de de-embedding (CDE). L'un est nommé « Thru », un autre nommé « Open » et les deux autres nommés « Short ». Une fois les mesures de ces dispositifs complémentaires réalisées, il est possible de déterminer une à une les impédances « parasites »  $G_i$  et  $Z_i$ . Ensuite nous pouvons soustraire leurs effets à la mesure de la cellule de test complète afin de déterminer uniquement les grandeurs intrinsèques du quadripôle DST. Cette méthode est répandue lors de caractérisation hyper-fréquences, toutefois le nombre élevé de cellule de de-embedding complémentaires nécessaires reste la contrainte majeure de cette méthode.

## II.1.2 Méthode de E.P. Vandamme dérivée et adaptée à nos contraintes

### II.1.2.a Etape de modélisation

Afin de limiter le nombre de cellules de de-embedding (CDE), trop coûteux en termes de surface, et du fait qu'un court circuit de qualité soit délicat à réaliser, nous proposons de modifier le modèle électrique représenté en Figure 2-4 en celui décrit en Figure 2-5. En effet la mesure des CDE « Thru » et « Open » suffit pour le calcul des éléments  $G_i$ ,  $Z_i$  de ce nouveau modèle. Pour parvenir à cette simplification on attribue un sens physique à ce modèle. Ceci n'est pas une obligation car il possède un caractère générique du fait de sa structure quadripolaire, mais dans ce cas la mesure des trois standards « open », « short » et « thru » reste nécessaire. Dans notre cas nous supposons que :

- L'admittance  $G_3$  modélise le couplage entre les deux ports de mesures,
- Les admittances en parallèle  $G_1$  et  $G_2$  modélisent le comportement des plots de contacts,
- Les impédances séries ( $Z_1$  et  $Z_2$ ) modélisent les pertes électriques dans les lignes d'accès et les admittances  $Y_3$  représentent le comportement capacitif ainsi que les éventuelles pertes diélectriques des lignes d'accès.

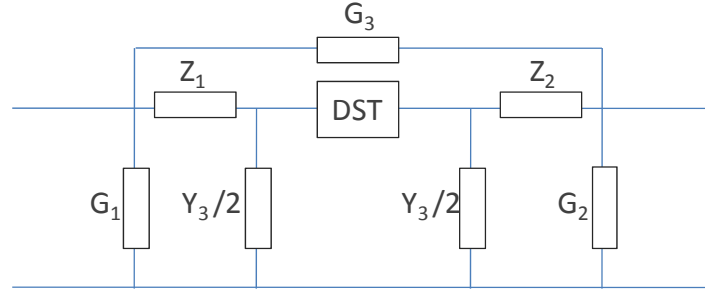


Figure 2-5 : Modèle électrique équivalent proposé pour la cellule de test (CDT) complète intégrant le dispositif de test (DST) à caractériser

### II.1.2.b Etape de détermination des éléments parasites $G_i$ , $Z_i$ , $Y_i$

#### Phase 1

Cette première phase consiste en la détermination des éléments  $G_1$  et  $G_2$  à l'aide à la CDE nommée OPEN. Cette CDE est réalisée en mesurant les plots de contact seul (Figure 2-6).

Le schéma électrique équivalent de cette CDE OPEN est un modèle en Pi, schématisé en Figure 2-7,

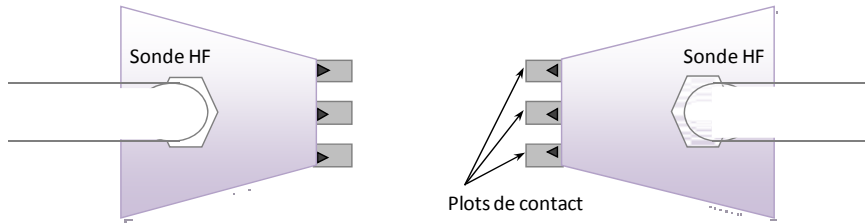


Figure 2-6 : Mesure de la cellule de test (CDE) OPEN.

La matrice admittance  $Y$  de la CDE « OPEN » est déduite de la mesure avec un analyseur de réseau vectoriel de sa matrice de diffusion  $S$  à l'aide des habituelles relations de passage :

$S \rightarrow Y$ , Équation 2-1 à Équation 2-4.

$$Y_{11} = \frac{1}{Z_0} \frac{(1 - S_{11})(1 + S_{22}) + S_{12} \times S_{21}}{(1 + S_{11})(1 + S_{22}) - S_{12} \times S_{21}} \quad \text{Équation 2-1}$$

$$Y_{12} = \frac{1}{Z_0} \frac{-2 \times S_{12}}{(1 + S_{11})(1 + S_{22}) - S_{12} \times S_{21}} \quad \text{Équation 2-2}$$

$$Y_{21} = \frac{1}{Z_0} \frac{-2 \times S_{21}}{(1 + S_{11})(1 + S_{22}) - S_{12} \times S_{21}} \quad \text{Équation 2-3}$$

$$Y_{22} = \frac{1}{Z_0} \frac{(1 + S_{11})(1 - S_{22}) + S_{12} \times S_{21}}{(1 + S_{11})(1 + S_{22}) - S_{12} \times S_{21}} \quad \text{Équation 2-4}$$

où  $Z_0$  est l'impédance de référence obtenue après le calibrage préliminaire de l'analyseur de réseau (généralement 50 Ohms)

Une identification des paramètres  $Y_{open\ ij}$  de la matrice admittance  $Y_{open}$  de la CDE OPEN avec les éléments  $G_i$  de son modèle équivalent (Équation 2-5 à Équation 2-7) permet d'obtenir ces derniers :

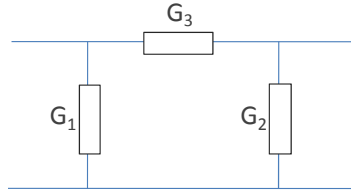


Figure 2-7 : Modèle électrique équivalent de la CDE OPEN

$$G_1 = Y_{Open\ 11} + Y_{Open\ 12} \quad \text{Équation 2-5}$$

$$G_2 = Y_{Open\ 22} + Y_{Open\ 21} \quad \text{Équation 2-6}$$

$$G_3 = -Y_{Open\ 21} \quad \text{Équation 2-7}$$

### Phase 2

Cette deuxième phase consiste en la détermination des éléments  $Z_1$ ,  $Z_2$ ,  $Y_3$  à l'aide de la CDE nommée THRU. Cette CDE complémentaire est réalisée en mesurant les plots de contact accolé aux lignes d'accès.

Le schéma électrique équivalent de la CDE THRU est un modèle en T, schématisé en Figure 2-8.

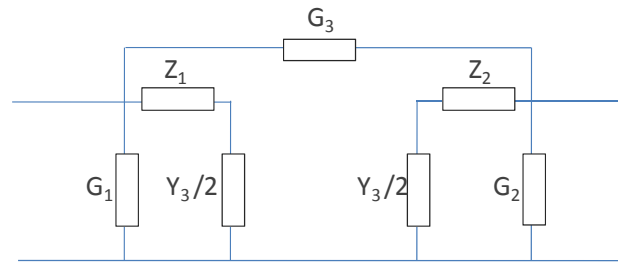


Figure 2-8 : Modèle électrique équivalent de la CDE THRU

Les paramètres  $S$  mesurés sur la CDE THRU sont transformés en paramètres  $Y$  (notée  $Y_{THRU}$ ). La contribution des plots de contact peut être ensuite retranchée puisque les admittances  $G_1$  et  $G_2$  ont précédemment été déterminées :

$$Y'_{THRU} = Y_{THRU} - \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \quad \text{Équation 2-8}$$

La matrice impédance (notée  $Z'_{THRU}$ ) du quadripôle représenté en Figure 2-10 est ensuite calculée grâce aux relations de passage :  $Y \rightarrow Z$ , Équation 2-9 à Équation 2-12.



$$Z_{11} = \frac{Y_{22}}{Y_{11} \times Y_{22} - Y_{12} \times Y_{21}} \quad \text{Équation 2-9}$$

$$Z_{12} = \frac{-Y_{12}}{Y_{11} \times Y_{22} - Y_{12} \times Y_{21}} \quad \text{Équation 2-10}$$

$$Z_{21} = \frac{-Y_{21}}{Y_{11} \times Y_{22} - Y_{12} \times Y_{21}} \quad \text{Équation 2-11}$$

$$Z_{22} = \frac{Y_{11}}{Y_{11} \times Y_{22} - Y_{12} \times Y_{21}} \quad \text{Équation 2-12}$$

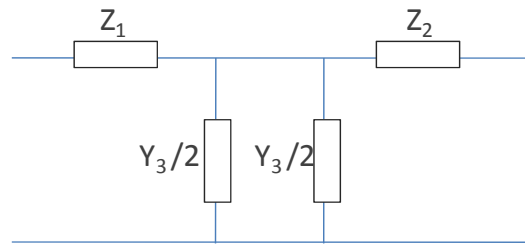


Figure 2-9 : Modèle électrique équivalent de la CDE THRU une fois les effets des plots de contact corrigés.

Les éléments  $Z_1$ ,  $Z_2$  et  $Y_3$  peuvent donc maintenant être déterminés par identification avec la matrice  $Z'$  obtenue précédemment : Équation 2-13 à Équation 2-15.

$$Z_1 = Z'_{\text{THRU11}} - Z'_{\text{THRU12}} \quad \text{Équation 2-13}$$

$$Z_2 = Z'_{\text{THRU22}} - Z'_{\text{THRU21}} \quad \text{Équation 2-14}$$

$$Y_3 = \frac{1}{Z'_{\text{THRU 21}}} \quad \text{Équation 2-15}$$

A la suite de ces deux phases l'ensemble des paramètres électriques « parasites »  $G_i$ ,  $Z_i$ ,  $Y_i$  du modèle de la cellule de test (CDE) est connu.

### II.1.2.c Etape d'épluchage ou de-embedding du DST

Il est maintenant possible d'extraire les caractéristiques intrinsèques du DST enfoui dans la cellule de test (CDT). La matrice  $Y_{\text{DST}}$  du dispositif sous test est extraite à partir de la matrice  $Y_{\text{mesure}}$  issue de la mesure sur la cellule de test complète, à laquelle sont retranchés les effets des différents éléments parasites  $G_i$ ,  $Z_i$  et  $Y_i$  présent dans le modèle électrique équivalent à cette cellule de test.

La relation utilisée s'exprime en terme de matrice admittance

$$Y_{DST} = \left( \left( Y_{mesure} - \begin{bmatrix} G_1 + G_3 & -G_3 \\ -G_3 & G_2 + G_3 \end{bmatrix} \right)^{-1} - \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \right)^{-1} - \begin{bmatrix} 1/2Z_1 & 0 \\ 0 & 1/2Z_3 \end{bmatrix} \quad \text{Équation 2-16}$$

#### II.1.2.d Conclusion

Cette méthode de de-embedding présente l'avantage d'être applicable à tout type de dispositif à caractériser, composants actifs ou passifs, localisés (condensateurs, inductances, TSV...) ou répartis comme les lignes de transmission que sont les interconnexions. Ses faiblesses résident dans les hypothèses faites au niveau du modèle représentant les effets parasites liés à l'environnement du dispositif sous test et le lien effectué avec chaque élément d'accès à ce DST. En effet modéliser les plots de contact par une admittance parallèle et les lignes d'accès au DST par deux impédances (série et parallèle) ne peut être parfaitement rigoureux. Néanmoins nous montrerons dans la dernière partie de ce chapitre que cette méthode est robuste et conduit à des résultats d'extraction tout aussi précis que les autres méthodes existantes.

## II.2 Méthode de De-Embedding destinées à la caractérisation de l'exposant de propagation d'une ligne de transmission

Le de-embedding d'une ligne de transmission enfouie dans un circuit revient à extraire avec précision son exposant de propagation  $\gamma$  et son impédance caractéristique  $Z_c$ .

Outre la méthode exposée précédemment qui s'applique à tous types de dispositifs, ce de-embedding peut être réalisé à l'aide de plusieurs techniques. Dans cette partie, une méthode permettant d'obtenir précisément l'exposant de propagation d'une ligne de transmission est détaillée. Dans le paragraphe II.3 quatre méthodes permettant l'extraction de l'impédance caractéristique seront présentées et comparées.

Avant de présenter la méthode d'épluchage appliquée à la caractérisation de l'exposant de propagation d'une ligne de transmission nous rappelons les principaux concepts d'analyse et de modélisation utilisés pour ces dernières

### II.2.1 Modèle des lignes de transmission

L'objectif de ce paragraphe est de rappeler brièvement la théorie des lignes de transmission, et plus particulièrement les grandeurs et le modèle électrique utilisés pour représenter les effets de

propagation sur ces lignes. Ce modèle sera celui utilisé pour la caractérisation des briques élémentaires de type interconnexion de RDL et interconnexion du BEOL.

Une ligne de transmission, propageant une onde TEM dans la direction  $z$  (Figure 2-10-a), peut être représentée par un modèle électrique équivalent. Ce schéma équivalent fait alors intervenir des selfs inductances, des capacitances, des résistances et des conductances. Pour un élément de longueur infinitésimale  $dz$  d'une ligne de transmission, on obtient le modèle présenté sur la Figure 2-10-b.

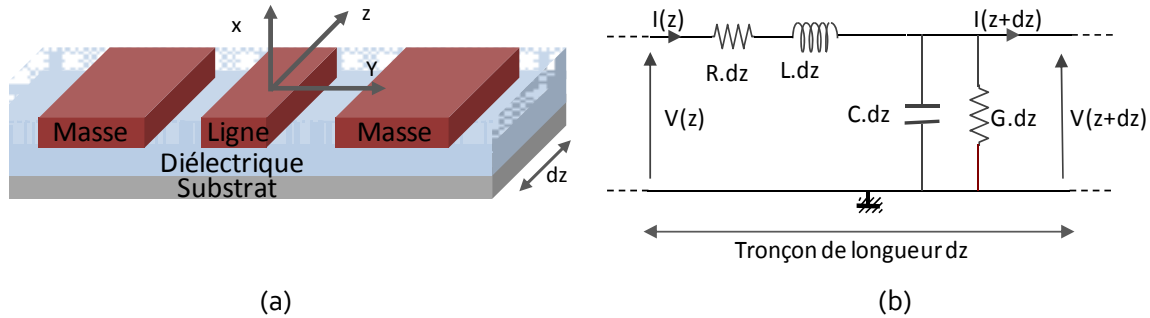


Figure 2-10 : (a) Tronçon de ligne de transmission. (b) Modèle électrique équivalent d'une ligne de transmission.

Ce modèle est un réseau qui comporte :

- en série, une résistance  $R$  en  $\Omega/m$  et une inductance  $L$  en  $H/m$ , représentant respectivement les pertes d'énergie active et réactive, à la fois dans les conducteurs et les matériaux magnétiques composant la ligne. Ces deux paramètres sont relatifs à la présence du champ magnétique qui s'établit lors de l'apparition d'un courant dans les conducteurs.

- en parallèle, une conductance  $G$  en  $S/m$  et une capacité  $C$  en  $F/m$  représentant respectivement les pertes d'énergie active et réactive dans les diélectriques composant la ligne. Ces deux paramètres sont relatifs à la présence du champ électrique qui s'établit lors de l'apparition d'une tension (ou DDP : différence de potentiel) entre les conducteurs.

On notera que ce modèle n'a de sens que si les champs électromagnétiques propagés (champs électrique et magnétique) sont dits TEM, c'est-à-dire si leurs orientations sont transversales par rapport au sens de propagation.

Le modèle électrique d'une ligne de transmission sera donc constitué de l'association d'une grande quantité de cette cellule de base.  $R$ ,  $L$ ,  $G$  et  $C$  sont les paramètres primaires linéiques de la ligne de transmission. Ce sont des grandeurs directement issues de la conformation physique de la ligne, ainsi que des caractéristiques des matériaux qui la composent. Par exemple, les paramètres  $C$  et  $G$  sont en partie liés respectivement à la permittivité  $\epsilon_r$  et aux pertes  $\epsilon''$  des isolants. Les paramètres primaires sont généralement dépendants de la fréquence.

Deux autres grandeurs permettent aussi de caractériser et de définir idéalement une ligne de transmission :

- l'exposant de propagation  $\gamma$  ( $\text{m}^{-1}$ )
- l'impédance caractéristique  $Z_c$  ( $\Omega$ )

Ces deux grandeurs sont nommées paramètres secondaires de la ligne. Ils s'expriment en fonction des paramètres primaires ( $R$ ,  $L$ ,  $G$  et  $C$ ) du modèle électrique de la ligne de transmission. Ils permettent de mettre en évidence des ondes de tension et de courant sur cette dernière. Nous rappelons que l'exposant de propagation  $\gamma$  traduit le comportement des signaux le long de l'interconnexion. C'est la grandeur complexe définie par :

$$\gamma = \alpha + j\beta \quad \text{Équation 2-17}$$

La partie réelle  $\alpha$  en Np/m exprime l'affaiblissement des signaux par unité de longueur. La partie imaginaire  $\beta$  en Rad/m, quant à elle, représente la rotation de phase des signaux par unité de longueur (par conséquent, dans le domaine temporel, leur retard de propagation le long de la ligne). Généralement l'exposant de phase  $\beta$  est relié à la vitesse de phase  $v$  par :  $\beta = \omega/v$  où  $\omega$  est la pulsation des ondes. Dès lors que l'exposant de propagation n'est pas purement imaginaire et que la vitesse de propagation dépend de la fréquence ( $\beta$  non proportionnel à  $\omega$ ) on parle de guide dispersif.

L'impédance caractéristique  $Z_c$  est définie comme étant l'impédance que l'on voit en entrée de ligne si cette dernière était infinie (onde progressive uniquement).

$\gamma$  et  $Z_c$  sont reliés aux paramètres primaires ( $R$ ,  $L$ ,  $G$  et  $C$ ) par les relations suivantes :

$$\gamma(\omega) = \sqrt{(R(\omega) + jL(\omega) \cdot \omega) \cdot (G(\omega) + jC(\omega) \cdot \omega)} = \alpha + j\beta \quad \text{Équation 2-18}$$

$$Z_c(\omega) = \sqrt{\frac{(R(\omega) + jL(\omega) \cdot \omega)}{(G(\omega) + jC(\omega) \cdot \omega)}} = \sqrt{\frac{Z(\omega)}{Y(\omega)}} \quad \text{Équation 2-19}$$

Avec  $\omega$  la pulsation des signaux en rad/s,  $Z(\omega)$  étant l'impédance série de la ligne de transmission due aux paramètres  $R$  et  $L$  et  $Y(\omega)$  l'impédance parallèle de la ligne de transmission due aux paramètres  $G$  et  $C$ .

## II.2.2 Extraction de l'exposant de propagation $\gamma$

Le calcul de l'exposant de propagation est indispensable à la caractérisation des lignes de transmission. Dans cette partie nous explicitons une façon simple et précise de l'obtenir.

A partir de la mesure de deux lignes de longueurs différentes  $L_1$  et  $L_2$  nous donnant respectivement les matrices de diffusion  $[S_{L_1}]$  et  $[S_{L_2}]$ , on extrait cet exposant de propagation  $\gamma$

composé d'une partie réelle  $\alpha$  (Np/m) représentant la constante d'atténuation et d'une partie imaginaire  $\beta$  (rad/m) représentant la constante de phase.

Les matrices  $[S_{L1}]$  et  $[S_{L2}]$  nous permettent de déterminer les matrices  $[T_{L1}]$  et  $[T_{L2}]$  grâce aux relations de passage S→T :

$$[T_{11}] = \frac{(-S_{11}S_{22} + S_{12}S_{21})}{S_{21}} \quad \text{Équation 2-20}$$

$$[T_{12}] = \frac{S_{11}}{S_{21}} \quad \text{Équation 2-21}$$

$$[T_{21}] = \frac{-S_{22}}{S_{21}} \quad \text{Équation 2-22}$$

$$[T_{22}] = \frac{1}{S_{21}} \quad \text{Équation 2-23}$$

Des propriétés matricielles (matrices semblables ayant le même déterminant et la même trace) nous permettent d'écrire que le produit matriciel  $[T_{L2}].[T_{L1}]$  est relié à  $\gamma$  via la relation :

$$[T_{L2}].[T_{L1}]^{-1} = \begin{bmatrix} e^{-\gamma|L_2 - L_1|} & 0 \\ 0 & e^{-\gamma|L_2 - L_1|} \end{bmatrix} \quad \text{Équation 2-24}$$

L'exposant de propagation peut donc être extrait grâce à la formule suivante :

$$\gamma = \frac{1}{|L_2 - L_1|} \operatorname{ach} \left( \frac{\operatorname{trace}([T_{L2}].[T_{L1}]^{-1})}{2} \right) \quad \text{Équation 2-25}$$

Par la suite cette méthode sera utilisée dès que l'on aura besoin de connaître l'exposant de propagation d'une ligne de transmission enfouie dans un circuit.

### II.3 Méthode de De-Embedding destinées à la caractérisation de l'impédance caractéristique d'une ligne de transmission

Déterminer expérimentalement l'impédance caractéristique d'une ligne de transmission est plus complexe et difficile qu'extraire son exposant de propagation puisque un analyseur de réseau travail en impédance normalisée ou réduite alors qu'une détermination d'impédance caractéristique nécessite la connaissance d'une référence d'impédance absolue. Il en résulte une problématique de détermination d'une référence d'impédance, difficile à établir dès lors que nous travaillons sur des dispositifs enfouis.

Dans cette partie quatre méthodes sont présentées pour extraire l'impédance caractéristique d'une ligne de transmission enfouie dans un circuit.

La première est basée sur les travaux de D.F. Williams [Williams, 1995] et nécessite de modéliser la transition entre les plots de contact sur la cellule de test et la ligne de transmission par un quadripôle particulier. La seconde méthode, nommée DUMMC (De-embedding Using Modification of Materials Characteristics) a été développée au laboratoire [Fourneaud-1, 2011] et elle est basée sur la modification des propriétés électriques des diélectriques utilisés dans la fabrication de la ligne de transmission. La troisième développée aussi au laboratoire [De-Rivaz, 2011] s'appuie sur la mesure de dispositifs complémentaires spécifiques en utilisant deux calibrages, OSTL et TRL. La dernière repose sur les travaux de T.M.Winkel [Winkel-2, 1996] que nous avons reformulés [Roullard-1, 2011] afin de contourner certaines hypothèses simplificatrices.

### II.3.1 Méthode basée sur les travaux de D.F. Williams

Cette méthode permet d'obtenir l'impédance caractéristique de lignes de transmission en modélisant la transition entre les plots de contact et la ligne de transmission sous test par un schéma électrique équivalent spécifique. Ce schéma électrique permet de faire le lien entre l'impédance de référence connue après un calibrage de type OSTL au niveau des sondes de mesure et l'impédance caractéristique que l'on cherche à déterminer. Nous schématisons le problème posé dans la Figure 2-11.

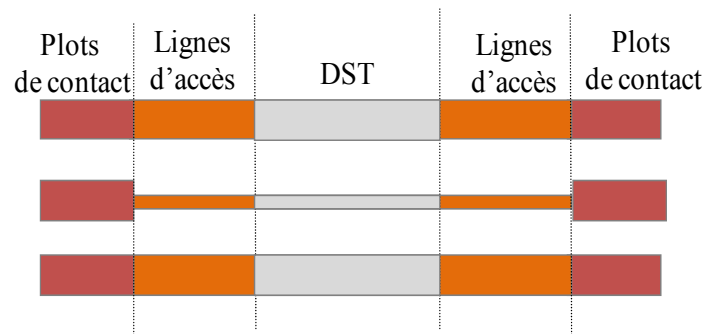


Figure 2-11 : CDT (plots+lignes accès +ligne sous test DST)

Les matrices de transfert T et l'impédance caractéristique recherchée est obtenue à l'aide d'une série de mesures effectuées sur les cellules de de-embedding (CDE) complémentaires donnés en Figure 2-12 et nommés Thru, Line, Reflect, sont utilisés.

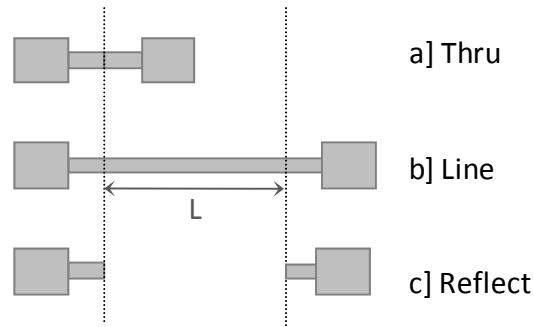


Figure 2-12 : CDE cellules de de-embedding nécessaire au calcul de l'impédance caractéristique.

La CDE Thru est défini comme étant la connexion directe les deux quadripôles de transition permettant l'accès au DST, ligne de transmission, dont nous recherchons l'impédance caractéristique. La CDE Line est défini comme étant la combinaison de ces deux quadripôles de transition et d'un tronçon de ligne de longueur  $L$ . La CDE « Reflect » est défini comme la mise en circuit ouvert des deux transitions d'accès et donc de fort coefficient de réflexion sur chacun des ports.

### II.3.1.a Étape de modélisation

Le modèle électrique équivalent de la transition entre le plot et la ligne d'accès au DST (Figure 2-13) se compose d'un élément admittance parallèle  $Y_P$  suivi d'un transformateur d'impédance. L'admittance représente le plot de contact avec des pertes modélisées sous forme d'une conductance  $G$  et une capacité en parallèle  $C$ . On a donc  $Y_P = G + j\omega C$ . Le transformateur d'impédance modélise le passage entre l'impédance de référence  $Z_R = 50\Omega$  et l'impédance que l'on cherche  $Z_C$ .

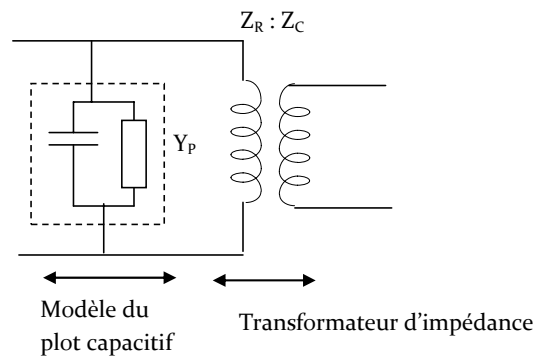


Figure 2-13 : Schéma électrique la transition entre le plot et la ligne d'accès au DST

### II.3.1.b Etape de détermination des termes d'erreur

*Phase 1 : Expression de la matrice de transfert de l'ensemble constitué par le plot de contact et la ligne d'accès au DST*

Les matrices  $[T]$  sont obtenues à partir des matrices  $[S]$  par la relation suivante :

$$[T] = \frac{1}{S_{21}} \begin{bmatrix} -\Delta & S_{11} \\ -S_{22} & 1 \end{bmatrix} \quad \Delta = S_{11}S_{22} - S_{12}S_{21} \quad \text{Équation 2-26}$$

La matrice de transfert  $[T]$  du modèle du plot de contact que l'on nomme  $[T_{\text{plot}}]$  s'écrit :

$$[T_{\text{plot}}] = \begin{bmatrix} \frac{2 - Y_p Z_R}{2} & -\frac{Y_p Z_R}{2} \\ \frac{Y_p Z_R}{2} & \frac{2 + Y_p Z_R}{2} \end{bmatrix} \quad \text{Équation 2-27}$$

La matrice de transfert  $[T]$  du transformateur d'impédance  $[T_{\text{transfo}}]$  s'écrit :

$$[T_{\text{transfo}}] = \begin{bmatrix} \frac{1}{\sqrt{1-\Gamma^2}} & \frac{\Gamma}{\sqrt{1-\Gamma^2}} \\ \frac{\Gamma}{\sqrt{1-\Gamma^2}} & \frac{1}{\sqrt{1-\Gamma^2}} \end{bmatrix} \quad \text{Équation 2-28}$$

Avec  $\Gamma$  coefficient de réflexion :

$$\Gamma = \frac{Z_R - Z_C}{Z_R + Z_C} \quad \text{Équation 2-29}$$

La matrice de transfert de l'ensemble plot de contact et ligne d'accès au DST nommée  $[T_s]$  s'écrit :

$$[T_s] = \frac{1}{\sqrt{1-\Gamma^2}} \left( \begin{bmatrix} 1 & \Gamma \\ \Gamma & 1 \end{bmatrix} + (1+\Gamma) \frac{Y_p Z_R}{2} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \right) \quad \text{Équation 2-30}$$

Phase 2 : Expression théorique de  $G$  et  $Z_c$  à partir des termes de la matrice de transfert  $T_s$

A l'issue de l'étape précédente, nous connaissons l'expression de la matrice de transfert de la zone d'accès au DST (plot + accès). Il s'agit maintenant de remonter à celle de l'impédance caractéristique du DST.

A partir de la relation précédente (Équation 2-30) et de deux paramètres ( $T_{S21}$  et  $T_{S12}$ ) de la matrice  $T_s$  le calcul de l'impédance caractéristique est possible.

Le terme  $T_{S21}$  de la matrice  $T_s$  est multiplié par :  $\frac{Y_p Z_r}{2}$ .



Le terme  $T_{s_{12}}$  de la matrice  $T_s$  est multiplié par :  $-\frac{Y_p Z_r}{2}$ .

Il suffit d'additionner ces deux termes pour éliminer l'admittance  $Y_p$ . Le coefficient de réflexion peut être exprimé en fonction de deux termes de la matrice  $T_s$ .

$$\Gamma = \sqrt{\frac{(T_{s_{12}} + T_{s_{21}})^2}{4 + (T_{s_{12}} + T_{s_{21}})^2}} \quad \text{Équation 2-31}$$

L'expression de l'impédance caractéristique  $Z_c$  recherchée se déduit facilement de l'Équation 2-31 à l'aide de l'Équation 2-29.

Si nous obtenons par la mesure les valeurs des termes  $T_{s_{12}}$  et  $T_{s_{21}}$  nous aurons accès à la valeur expérimentale du coefficient de réflexion  $\Gamma$  et par conséquent à la valeur de l'impédance caractéristique  $Z_c$  de la ligne de transmission sous test en utilisant cette Équation 2-29.

Phase 3 : Extraction de l'impédance caractéristique de la ligne de transmission sous test

L'étape suivante est donc d'obtenir les valeurs expérimentales des paramètres  $[T_s]$  de la transition plot de contact et la ligne d'accès au DST. Cette procédure, qui fait appel à de lourdes équations, est présentée en annexe 2, elle permet de déterminer la matrice de transfert de la transition à partir des trois dispositifs de de-embedding Thru, Line et Reflect.

### II.3.1.c Conclusion

Cette approche présente l'avantage de ne pas calculer l'admittance  $Y_p$ . Ainsi la forme du plot de contact n'a pas d'influence sur la détermination de l'impédance caractéristique. Toutefois une des limites de cette approche réside dans la transition ligne / plot. En effet le modèle électrique proposé pour la transition est un modèle localisé. Le schéma électrique ne modélise pas d'éventuels phénomènes de propagation.

Le choix du signe obtenu pour le coefficient de réflexion reste toutefois une étape délicate. Il s'effectue en estimant la valeur de l'impédance caractéristique que l'on cherche à déterminer. Ainsi une valeur d'impédance supérieure à  $Z_c$  conduit à un signe positif.

### II.3.2 Méthode DUMMC (De-embedding using modification of materials characteristics).

Cette méthode est basée sur la possibilité de modifier des propriétés électriques des matériaux diélectriques intégrés dans l'architecture de la cellule sous test. On profite en effet d'une série de mesures, réalisées avec et sans perturbation sur ces matériaux diélectriques, pour obtenir un nombre  $N$  d'équations permettant de résoudre un système d'équations à  $n$  inconnues, avec  $N > n$ . Le principe de cette méthode de de-embedding a été développé au laboratoire [Fourneaud-1, 2011].

Son application nécessite les deux structures de test suivantes:

- La cellule de test (CDT) incluant la ligne de transmission sous test (DST) de longueur  $L_{DST}$ , insérée entre les plots de contact et les lignes d'accès au DST.
- La même cellule de test mais pour laquelle la ligne sous test est plus courte, de longueur  $L = L_{DST} - \Delta L$ . Cette structure de test peut être considérée comme un standard ou une cellule de de-embedding (CDE) et elle est le seul dispositif supplémentaire nécessaire.

#### II.3.2.a Etape de modélisation

La ligne de transmission sous test (DST) est caractérisée par son exposant de propagation  $\gamma$  et par son impédance caractéristique  $Z_c$  que l'on cherche à déterminer.

La Figure 2-14 schématise ces structures. Les quadripôles  $Q$  et  $Q'$  représentent le comportement de l'ensemble constitué des plots d'accès mis en cascade avec les lignes d'accès à la ligne sous test (DST). Nous les supposons identiques en émettant l'hypothèse de conditions d'accès au DST symétriques (ce qui est justifié si nous considérons un bon design et une bonne précision dans leur réalisation technologique). Par la suite nous supposerons  $Q = Q'$ .

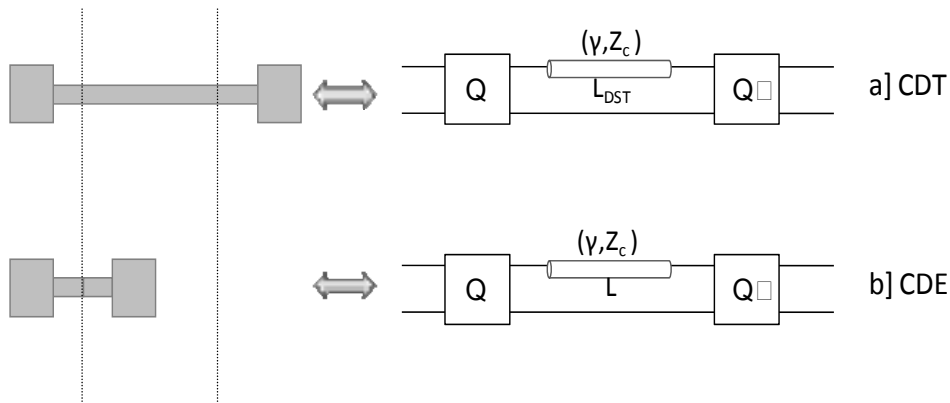


Figure 2-14 : Cellule de test (CDT) et cellule de de-embedding (CDE) nécessaire à l'application de la méthode DUMMC

## Phase 1 : Modélisation de la cellule de de-embedding

Le modèle électrique de la cellule de De-Embedding, présenté en Figure 2-15, est totalement générique puisqu'il s'agit d'un composant passif. Nous faisons le choix d'un modèle équivalent en T.

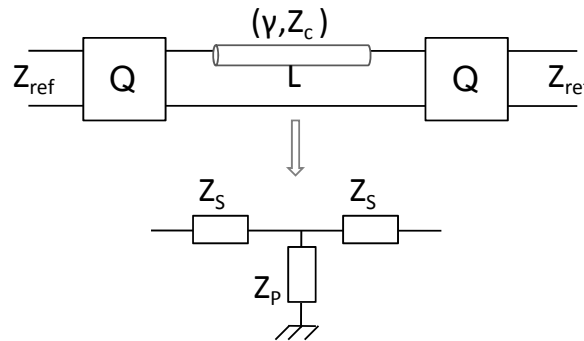


Figure 2-15 : Modèle électrique de la cellule de de-embedding (CDE),

La mesure de la matrice de diffusion de la CDE et sa transformation en matrice ABCD ( $Z_{ref}$  est connue après calibrage OSTL au niveau des sondes de mesures RF) permet aisément d'extraire les impédances  $Z_s$  et  $Z_p$ . Par la suite nous pouvons donc les considérer comme connues par mesure.

Ce modèle en T équivalent de la cellule de de-embedding CDE peut être décomposé selon l'association de deux sous modèles en T symétriques, correspondant chacun à la demi-structure et constitués des impédances  $Z_A$ ,  $Z_B$ ,  $Z_C$ .

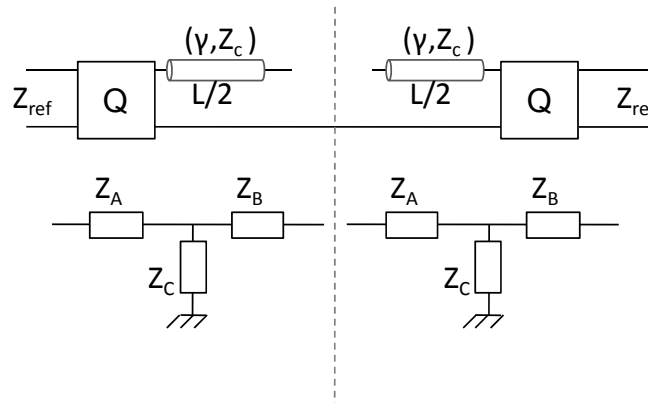


Figure 2-16 : Cellule de de-embedding (CDE) divisée en deux avec des sous modèles en T identiques.

Notre objectif est de déterminer  $Z_A$ ,  $Z_B$  et  $Z_C$  mais la mesure de la matrice de diffusion de la CDE ne suffit pas pour les extraire, car nous obtenons seulement deux équations pour déterminer trois inconnues.

C'est ici qu'intervient l'idée de modifier les caractéristiques électriques des diélectriques présents dans les structures de test, afin d'obtenir un nombre d'équations suffisant par rapport au nombre d'inconnues

Phase 2 : Modification des propriétés des diélectriques dans les cellules de test.

Nous supposons la possibilité de perturber les caractéristiques des diélectriques constituant la cellule de test et la cellule de de-embedding. Une des voies que nous avons utilisée est d'éclairer ces cellules avec un rayonnement électromagnétique de longueur d'onde située dans le visible afin de modifier les caractéristiques du substrat de silicium sur lequel nos lignes sont gravées.

Nous pouvons donc considérer deux types de mesure sur la même CDE : avec et sans lumière (Figure 2-17). Elles nous permettront ensuite de déterminer les impédances  $Z_A$ ,  $Z_B$  et  $Z_C$  des demi-structures présentées en Figure 2-16.

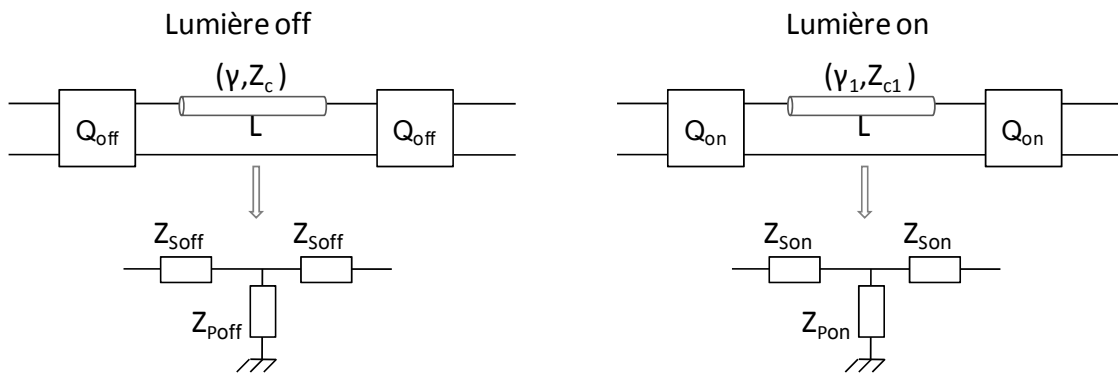


Figure 2-17 : Cellule de de-embedding(CDE) mesurée selon deux configurations: côté gauche, sans lumière (Off) et côté droite, avec lumière (On).

Dans le cas d'un mode de propagation TEM ou quasi-TEM, il est rationnel de proposer l'hypothèse suivante: le faisceau de lumière modifie seulement l'impédance parallèle  $Z_C$  de la demi-structure. En effet,  $Z_C$  exprime l'impédance entre les masses et le signal, elle dépend donc fortement du silicium dont les propriétés varient avec la lumière. En revanche  $Z_A$  et  $Z_B$  étant particulièrement liées aux conducteurs métalliques, l'effet de la lumière n'impacte pas leurs valeurs. Ainsi, les deux configurations proposées dans la Figure 2-18 sont obtenues, dans lesquelles les impédances  $Z_A$  et  $Z_B$  restent inchangées quelle que soit la présence ou non de lumière.

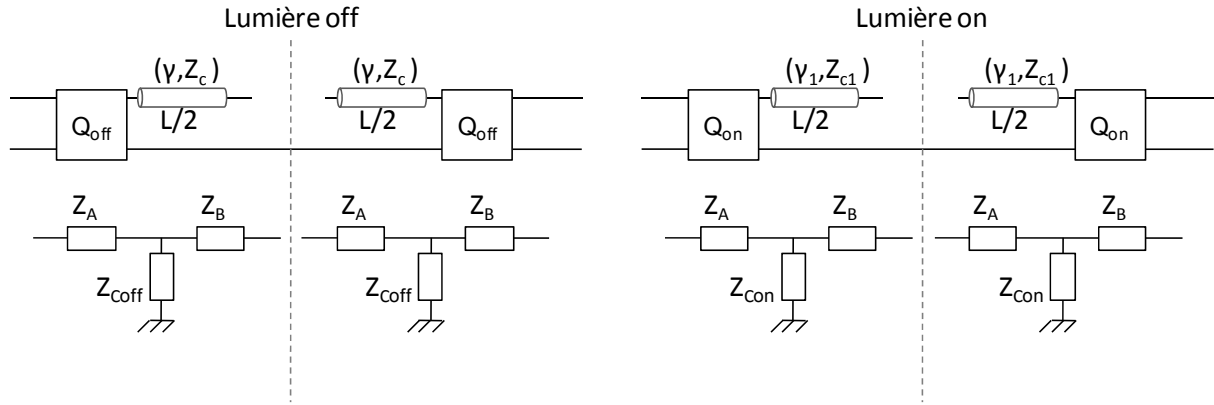


Figure 2-18 : Cellule de de-embedding (CDE) divisée en deux et mesurée selon deux configurations: côté gauche sans lumière (Off) et côté droit avec lumière (On).

De ces deux configurations (avec et sans lumière), il est maintenant possible d'établir un système d'équations (Équation 2-32 à Équation 2-35) permettant de déterminer les quatre impédances  $Z_A$ ,  $Z_B$  et  $Z_{con}$  et  $Z_{coff}$  de la demi-structure. L'indice Off représente une configuration sans lumière et l'indice On représente une configuration avec lumière. Il est intéressant de noter que  $Z_{Soff}$  (respectivement  $Z_{Son}$ ) et  $Z_{Poff}$  (respectivement  $Z_{Pon}$ ) sont connus grâce aux éléments de la matrice ABCD (Équation 2-36) obtenus à partir des mesures des paramètres S [Pozar, 2005].

$$Z_{Soff} = \frac{Z_B Z_{Coff} + Z_A (Z_B + Z_{Coff})}{Z_B + Z_{Coff}} \quad \text{Équation 2-32}$$

$$Z_{Poff} = \frac{Z_{Coff}^2}{2(Z_B + Z_{Coff})} \quad \text{Équation 2-33}$$

$$Z_{Son} = \frac{Z_B Z_{Con} + Z_A (Z_B + Z_{Con})}{Z_B + Z_{Con}} \quad \text{Équation 2-34}$$

$$Z_{Pon} = \frac{Z_{Con}^2}{2(Z_B + Z_{Con})} \quad \text{Équation 2-35}$$

$$Z_P = \frac{1}{C} \quad Z_S = \frac{(A-1)}{C} = \frac{(D-1)}{C} \quad \text{Équation 2-36}$$

Le système d'équations précédent nous donne, pour chaque point de fréquence, deux couples de quatre solutions ( $Z_A$ ,  $Z_B$ ,  $Z_{Coff}$  et  $Z_{Con}$ ). Le couple de bonne solution est déterminé en utilisant le critère suivant: la partie réelle de chaque impédance doit être positive.

Après la sélection du bon couple de solutions, une matrice ABCD de la demi-structure sans lumière peut être formée à partir des impédances  $Z_A$ ,  $Z_B$  et  $Z_{Coff}$  (Équation 2-37). La matrice ABCD des demies cellules de de-embedding, que l'on notera  $ABCD_{demi}$ , sera utilisée pour ce de-Embedding.

$$A_{demi} = 1 + \frac{Z_A}{Z_{Coff}}; \quad B_{demi} = Z_A + Z_B + \frac{Z_A Z_B}{Z_{Coff}}$$

$$C_{demi} = 1/Z_{Coff}; \quad D_{demi} = 1 + \frac{Z_B}{Z_{Coff}}$$

Équation 2-37

### II.3.2.b Etape d'épluchage ou de-embedding du DST

La cellule de test (CDT) composée avec les mêmes accès mais intégrant une ligne de transmission plus longue que dans la cellule de de-embedding (CDE) est également mesurée (Figure 2-14). Elle est constituée avec les mêmes éléments que la CDE, dont les paramètres électriques sont maintenant bien connus et d'un tronçon de ligne sous test (DST) inséré au milieu (Figure 2-19).

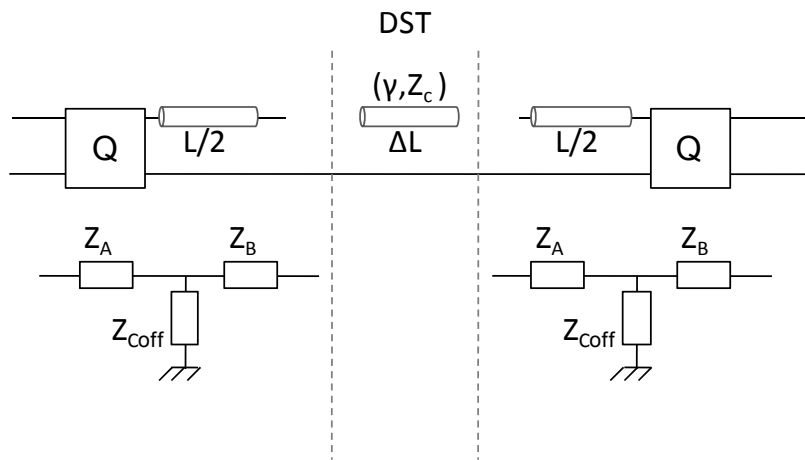


Figure 2-19 : Cellule de test composée des accès et d'un tronçon de ligne de transmission sous test (DST)

A partir de la mesure de la CDT et grâce à la théorie des matrices cascables ABCD, la matrice  $ABCD_{DST}$  de l'interconnexion sous test (DST) peut facilement être extraite en utilisant la formule suivante (Équation 2-38).

$$ABCD_{DST} = ABCD_{demi}^{-1} \cdot ABCD_{CDT} \cdot DBCA_{demi}^{-1}$$

Équation 2-38

$ABCD_{demi}$  représente la matrice ABCD de la demi-structure de la CDE et  $ABCD_{CDT}$  représente la matrice ABCD de la CDT. La matrice  $DBCA_{demi}$  représente la matrice ABCD de la demi-structure inversée, obtenue en inversant les paramètres A et D de la matrice  $ABCD_{demi}$ .

Au final nous pouvons extraire sans difficulté la constante de propagation de la ligne sous test et son impédance caractéristique à partir de la matrice  $ABCD_{DST}$  selon son expression donnée en (Équation 2-39).

$$ABCD_{DST} = \begin{bmatrix} \cosh(\gamma_0 \Delta L) & Z_{C0} \sinh(\gamma_0 \Delta L) \\ \frac{1}{Z_{C0}} \sinh(\gamma_0 \Delta L) & \cosh(\gamma_0 \Delta L) \end{bmatrix} \quad \text{Équation 2-39}$$

### II.3.2.c Conclusion

Sa puissance réside dans son caractère générique à la condition de travailler avec des diélectriques dont il est possible de modifier les caractéristiques électriques par le biais d'une perturbation extérieures (lumière, polarisation...). Sa faiblesse est située dans l'hypothèse que seule l'impédance parallèle, représentant la moitié de la CDE, est modifiée par cette perturbation sur les diélectriques. Cette hypothèse a été vérifiée et validée [Fourneaud-1, 2011] dans le cas où notre cellule de de-embedding (CDE) est constituée de plots de contact et d'un tronçon de ligne de transmission, cette impédance parallèle étant la seule représentative des isolants utilisés dans la structure.

## II.3.3 Méthode d'épluchage par bloc

### II.3.3.a Présentation générale

Cette méthode a aussi été développée au laboratoire [De Rivaz, 2011] et elle repose d'une part sur une décomposition de la cellule de test (CDT) présentée en Figure 2-20 en plusieurs blocs, chacun étant épluché et d'autre part sur l'utilisation des propriétés matricielles attribuables à des quadripôles passifs et donc réciproques. Elle nécessite plusieurs cellules de de-embedding (CDE), présentées en Figure 2-21 et dont les rôles sont précisés.

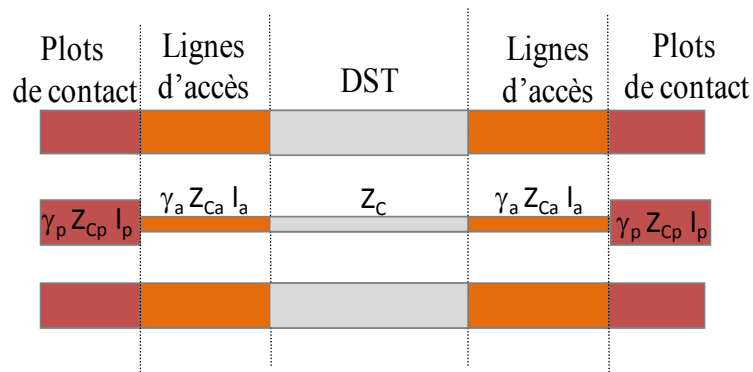


Figure 2-20 : CDT (plots+ lignes d'accès+DST).

Les CDE nommées « THRU ligne » et « LINE ligne » constituées de la cascade plot+ligne d'accès+plot comme illustré en Figure 2-21, sont mesurées après un calibrage de type OSTL en bout des sondes de mesure pour obtenir leur matrice [S].

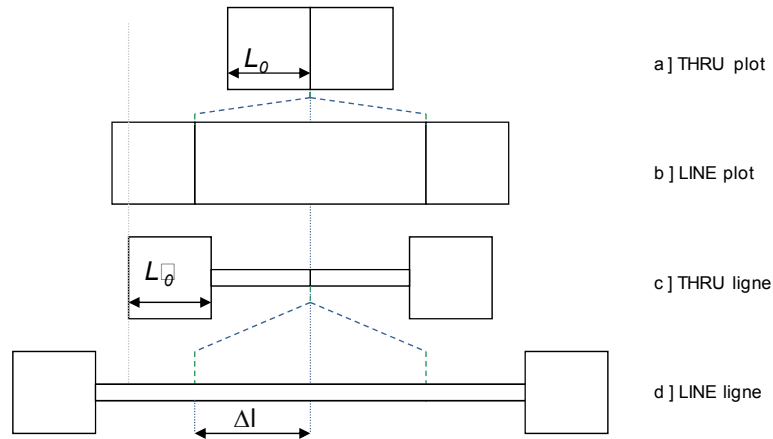


Figure 2-21 : Cellules de de-embedding nécessaires au calcul des exposants de propagation et des impédances caractéristiques des zones plot et ligne

Une seconde série de mesures complémentaires est effectuée sur les CDE nommées « THRU plot » et « LINE plot », il s'agit d'extraire le coefficient de réflexion  $\Gamma$  de la transition entre les plots de contact et la ligne d'accès. Ce dernier est extrait après un calibrage de type TRL avec les CDE associées aux plots. Ce calibrage TRL permet de placer les plans de références de mesure sur les plots. Il est alors possible d'en déduire le coefficient de réflexion  $\Gamma$  de la transition entre un plot la ligne en utilisant la première série de mesure sur la CDE « THRU line ».

Avec ces séries de mesures il est aussi très facile d'extraire les exposants de propagation  $\gamma_p$  et  $\gamma_a$  sur chaque zone plot et ligne d'accès présentes dans les CDE nommées « LINE plot » et « LINE Ligne ». Ensuite le choix de modèles électriques équivalents aux plots et aux lignes d'accès permet de déterminer les matrices ABCD de chaque zone, plot et ligne d'accès, constituant les ports d'accès à dispositif sous test (DST), comme rappelé en Figure 2-19. De ces matrices ABCD ont déduit facilement les impédances caractéristiques  $Z_{cp}$  et  $Z_{ca}$  de chaque zone, plot et ligne d'accès. Cette dernière étape permet de terminer la procédure de de-embedding et de calculer la matrice ABCD intrinsèque du DST enfoui dans la CDT (Figure 2-20).

Pour cela les hypothèses suivantes sont nécessaires :

- La transition entre le plot de contact et la ligne d'accès se comporte comme un saut d'impédance.
- La propagation des ondes sur tous les dispositifs doit être TEM.
- Les deux CDE « Thru plot » et « Thru line » sont tous deux réciproques et symétriques



### II.3.3.b Etape de mise en équations et de modélisation

La Figure 2-22 présente le modèle électrique pour chaque zone, la zone plot et la zone ligne d'accès sont tout simplement considérées comme des tronçons de ligne de transmission. Ces accès au dispositif à mesurer sont constitués de la cascade d'un plot de contact ( $Z_{cp}$ ,  $\gamma_p$ ,  $l_p$ ,  $l_p'$ ) et d'une portion de ligne de transmission ( $Z_{ca}$ ,  $\gamma_a$ ,  $l_a$ ). Une ligne de référence  $R$  symbolise l'endroit où sont posées les pointes de la sonde de mesure.

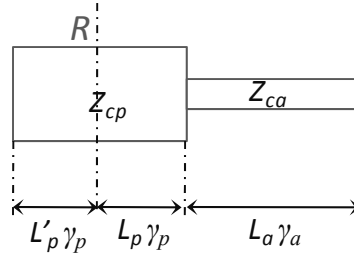


Figure 2-22 : Modèle électrique des zones plot et ligne de transmission

La partie de « stub » en circuit ouvert » à gauche de  $R$  est modélisée par une admittance  $Y_p'$  assimilable à un élément de propagation sur la partie ( $\gamma_p$ ,  $l_p'$ ) et un circuit ouvert ( $Z_L = \infty$ ), dans ce cas  $Y_p'$  s'écrit :

$$Y_p' = \frac{th(\gamma_p l_p')}{Z_{c_p}} \quad \text{Équation 2-40}$$

Remarque : Il est important de noter que le calcul de cette admittance ne sera pas nécessaire pour appliquer cette méthode de de-embedding.

$\Gamma$  est le coefficient de réflexion entre le plot et la ligne de transmission et il est donné par la relation suivante :

$$\Gamma = \frac{Z_{c_a} - Z_{c_p}}{Z_{c_a} + Z_{c_p}} \quad \text{soit} \quad \Gamma = \frac{\frac{Z_{c_a}}{Z_{c_p}} - 1}{\frac{Z_{c_a}}{Z_{c_p}} + 1} \quad \text{soit} \quad \frac{Z_{c_a}}{Z_{c_p}} = \frac{1 + \Gamma}{1 - \Gamma} \quad \text{Équation 2-41}$$

Dans la suite nous allons déterminer sa valeur.

La CDE « Thru line » peut être décomposée en deux parties symétriques contenant chacune un plot d'accès et une demi-longueur de ligne d'accès (Figure 2-23). Les modèles de la matrice chaîne  $abcd$  pour l'accès gauche et par réciprocité de la matrice  $dbca$  pour l'accès droit, sont les suivants :

$$abcd = \begin{bmatrix} a_0 & b_0 \\ c_0 + a_0 Y_p' & d_0 + b_0 Y_p' \end{bmatrix} \quad dbca = \begin{bmatrix} d_0 + b_0 Y_p' & b_0 \\ c_0 + a_0 Y_p' & a_0 \end{bmatrix} \quad \text{Équation 2-42}$$

Où  $a_0 b_0 c_0 d_0$  désigne la matrice chaîne de la cascade plot + ligne d'accès sans les stubs (Figure 2-24).

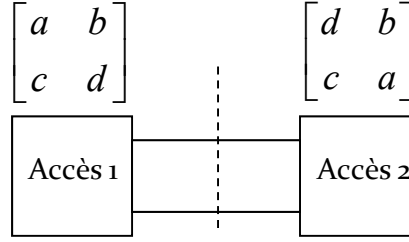


Figure 2-23 : Modélisation schématique des dispositifs d'accès gauche et droit de la CDE « Thru Line »

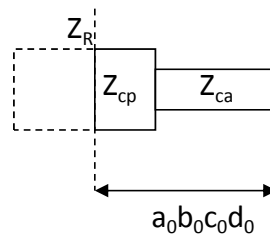


Figure 2-24 : Définition de la matrice  $a_0 b_0 c_0 d_0$

Conformément aux hypothèses de modélisation, on obtient les expressions des paramètres  $a_0 b_0 c_0 d_0$  :

$$a_0 = a = \cosh(\gamma_p l_p) \cosh(\gamma_a l_a) + \sinh(\gamma_p l_p) \sinh(\gamma_a l_a) \times \frac{1 - \Gamma}{1 + \Gamma} \quad \text{Équation 2-43}$$

$$d_0 = \cosh(\gamma_p l_p) \cosh(\gamma_a l_a) + \sinh(\gamma_p l_p) \sinh(\gamma_a l_a) \times \frac{1 + \Gamma}{1 - \Gamma} \quad \text{Équation 2-44}$$

$$b_0 = b = \left[ \cosh(\gamma_p l_p) \sinh(\gamma_a l_a) + \cosh(\gamma_a l_a) \sinh(\gamma_p l_p) \frac{1 - \Gamma}{1 + \Gamma} \right] \times Z_{ca} \quad \text{Équation 2-45}$$

$$c_0 = \left[ \cosh(\gamma_p l_p) \sinh(\gamma_a l_a) + \cosh(\gamma_a l_a) \sinh(\gamma_p l_p) \times \frac{1 + \Gamma}{1 - \Gamma} \right] / Z_{ca} \quad \text{Équation 2-46}$$

On constate que :

$a (= a_0)$  ne dépend que de  $(\gamma_p l_p)$ ,  $(\gamma_a l_a)$  et  $\Gamma$

$b (= b_0)$  dépend de  $(\gamma_p l_p)$ ,  $(\gamma_a l_a)$ ,  $\Gamma$  et de  $Z_{ca}$

### II.3.3.c Étapes de mesures

Un premier calibrage TRL est effectué avec les CDE « THRU plot » et « LINE plot », plaçant les plans de mesure à une distance  $L_0$  et fixant l'impédance de référence à l'impédance caractéristique  $Z_{cp}$  (qui reste inconnue). Puis dans un deuxième temps on mesure la CDE « Thru line » afin d'obtenir

le rapport  $Z_{ca} / Z_{cp}$  qui se calcule à partir des paramètres  $S$  mesurés sur la CDE « THRU Line » (Équation 2-47).

$$\frac{Z_{ca}}{Z_{cp}} = \sqrt{\frac{1 + S_{11} + S_{22} + S_{11}S_{22} - S_{12}S_{21}}{1 - S_{11} - S_{22} - S_{11}S_{22} - S_{12}S_{21}}} \Rightarrow \Gamma = \frac{Z_{ca}/Z_{cp} - 1}{Z_{ca}/Z_{cp} + 1} \quad \text{Équation 2-47}$$

$\Gamma$  est donc maintenant connu par voie expérimentale. Il en est de même pour les exposants de propagation  $\gamma_p$  et  $\gamma_a$ , déterminés avec ces séries de mesure et la méthode présentée au paragraphe II.2.2.

On cherche ensuite à extraire la valeur des éléments des matrices  $abcd$  et  $dcba$ , telles qu'elles sont présentées en Figure 2-23 et Équation 2-42. Pour cela on utilise encore les mesures de la matrice  $S$  de la CDE « TRHU ligne » et on en déduit sa matrice chaîne mesurée  $A_m B_m C_m D_m$ .

Si on note  $ABCD$  la matrice chaîne théorique de la CDE « Thru Ligne » complète et si on suppose que les demi-structures la composant sont réciproques ( $ad-cd=1$ ), cette matrice chaîne  $ABCD$  s'écrit :

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & b \\ c & a \end{bmatrix} = \begin{bmatrix} ad+bc & 2ab \\ 2cd & ad+bc \end{bmatrix} = \begin{bmatrix} 2ad-1 & 2ab \\ 2cd & 2ad-1 \end{bmatrix} \quad \text{Équation 2-48}$$

On peut alors identifier cette matrice chaîne de la CDE « Thru Ligne » à celle mesurée ( $A_m B_m C_m D_m$ ) avec l'Équation 2-49:

$$\begin{bmatrix} A_m & B_m \\ C_m & D_m \end{bmatrix} = \begin{bmatrix} 2ad-1 & 2ab \\ 2cd & 2ad-1 \end{bmatrix} \quad \text{Équation 2-49}$$

Or le paramètre  $a$  exprimé en Équation 2-48 est connu car calculé à partir de la connaissance de  $(\gamma_p l_p), (\gamma_a l_a), \Gamma$ .

De l'Équation 2-49 on en déduit alors  $b, c$  et  $d$  :

$$b = \frac{1}{2} \frac{B_m}{a} \quad d = \frac{A_m + 1}{2a} \quad \text{et} \quad c = \frac{C_m}{2d} \quad \text{Équation 2-50}$$

L'Équation 2-50 nous permet d'écrire  $B_m$ , à partir des Équation 2-43 et Équation 2-45, en fonction des paramètres de la ligne et du plot (Équation 2-51).

$$B_m = 2 \left( \cosh(\gamma_p l_p) \cosh(\gamma_a l_a) + \sinh(\gamma_p l_p) \sinh(\gamma_a l_a) \times \frac{1-\Gamma}{1+\Gamma} \right) \left( \cosh(\gamma_p l_p) \sinh(\gamma_a l_a) + \cosh(\gamma_a l_a) \sinh(\gamma_p l_p) \frac{1-\Gamma}{1+\Gamma} \right) \times Z_{ca}$$

$$\text{Équation 2-51}$$

$B_m$  étant connu par mesure, seule  $Z_{ca}$  reste inconnue, à partir de l'Équation 2-51, l'impédance caractéristique de la ligne d'accès  $Z_{ca}$  est donc déterminée avec l'Équation 2-52:

$$Z_{c_l} = Z_{c_a} = Z_0 \times \frac{(1 + S_{11})(1 + S_{22}) - S_{12}S_{21}}{4S_{21} \times \left( \cosh(\gamma_p l_p) \sinh(\gamma_a l_a) + \cosh(\gamma_a l_a) \sinh(\gamma_p l_p) \frac{1 - \Gamma}{1 + \Gamma} \right) \left( \cosh(\gamma_p l_p) \cosh(\gamma_a l_a) + \sinh(\gamma_p l_p) \sinh(\gamma_a l_a) \frac{1 - \Gamma}{1 + \Gamma} \right)}$$

Équation 2-52

Où  $B_m$  a été transformé par son expression avec les paramètres S mesurés sur la CDE « THRU Line » et l'impédance  $Z_0$  (50 Ohms) de référence issue du calibrage OSTL initiale en bout des sondes.

$Z_{ca}$  étant maintenant connue il est possible de déterminer sans difficulté la matrice ABCD intrinsèque d'un DST enfoui dans la cellule de test (CDT) et ainsi terminer l'opération de de-embedding.

### II.3.3.d Conclusion

Cette méthode de de-embedding est rigoureuse mais sa mise en œuvre est un peu plus lourde que les précédentes du fait du nombre plus élevé de cellules de de-embedding à fabriquer et par conséquent du nombre plus élevé de mesure à réaliser. Sa seule faiblesse reste dans l'hypothèse que la transition entre un plot de contact et une ligne de transmission se modélise par un seul saut d'impédance. Cette hypothèse a néanmoins été vérifiée lors de campagne de mesure, en effet la valeur du coefficient de réflexion  $\Gamma$  obtenu par mesure reste toujours extrêmement proche de la valeur obtenue par une modélisation fullwave avec HFSS de la structure étudiée.

## II.3.4 Méthode de de-embedding de TM. Winkel améliorée

### II.3.4.a Présentation générale

Dans ses travaux T.M. Winkel [Winkel-1, 1996] présente une technique de détermination de l'impédance caractéristique d'une interconnexion enfouie entre des plots d'accès (Figure 2-25). Nous avons reformulé son approche afin d'éviter des hypothèses simplificatrices [Roullard-1, 2011].

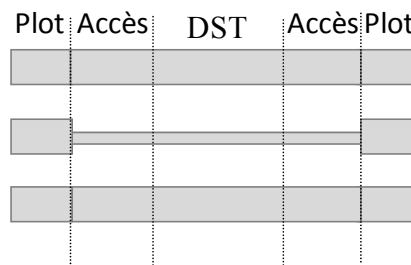


Figure 2-25 : CDT (plots+ lignes d'accès+DST).

Nous montrons qu'à partir de deux familles de cellules de de-embedding (Figure 2-26), on peut remonter à la valeur de l'impédance caractéristique  $Z_c$  de la ligne d'accès, que l'on considère ici comme notre dispositif sous test (DST) sans utiliser les simplifications induites par un modèle électrique équivalent. L'extraction de  $Z_c$  nécessite une famille de CDE correspondante aux plots de contact (Figure 2-26.a), qui comprend des structures de transmission de largeur égale à celle des plots, et d'une famille pour la partie ligne de transmission sous test (Figure 2-26.b). Ainsi il est nécessaire d'effectuer plusieurs séries de mesures de paramètres  $S_{ij}$ , pour des longueurs  $L$  différentes, après un calibrage OSTL au niveau des sondes de mesures.

Nous traiterons séparément les éléments issus des mesures des CDE de type plot et ceux issus du des CDE de type ligne. Cela nous permet dans un premier temps de déterminer l'impédance caractéristique des plots  $Z_p$ , puis l'impédance caractéristique de la ligne  $Z_c$ .

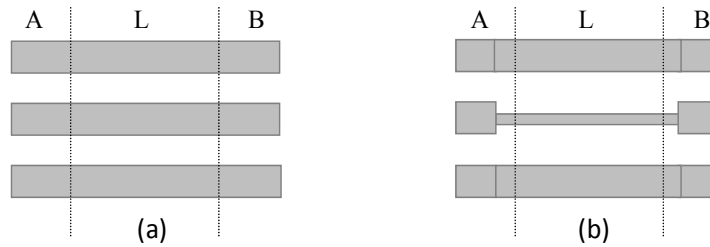


Figure 2-26 : Cellules de de-embedding, vue de dessus.

(a) Structure pour la détermination de l'impédance caractéristique des plots.

(b) Structure pour la détermination de l'impédance caractéristique de l'interconnexion sous test.

### II.3.4.b Mise en équations

Dans le but de déterminer l'impédance caractéristique des plots de contacts, deux CDE de la famille plots de longueurs différentes notées  $L_1$  et  $L_2$  sont mesurées. Pour ces longueurs  $L_1$  et  $L_2$  les matrices  $S_{ij}$  de mesures sont transformées en matrices  $[T_{m1}]$  et  $[T_{m2}]$  de transfert cascables mesurées (indice m). Ainsi :

$$[T_{m1}] = [T_A] \cdot [T_{L1}] \cdot [T_B] \quad \text{Équation 2-53}$$

$$[T_{m2}] = [T_A] \cdot [T_{L2}] \cdot [T_B] \quad \text{Équation 2-54}$$

$[T_A]$ ,  $[T_L]$ ,  $[T_B]$  correspondent respectivement aux matrices  $[T]$  des zones [A], [L] et [B].

Il est possible d'éliminer  $[T_B]$ , par l'opération :

$$[T_{m2}] \cdot [T_{m1}]^{-1} = [T_A] \cdot [T_{L2}] \cdot [T_{L1}]^{-1} \cdot [T_A]^{-1} \quad \text{Équation 2-55}$$

On pose :

$$[T_{m2}] \cdot [T_{m1}]^{-1} = [T_{mp}] \quad \text{Équation 2-56}$$

En transformant la matrice  $[T_{mp}]$  en matrice ABCD, notée  $[A_{mp}]$ , il est possible de calculer la valeur d'un coefficient intermédiaire, noté XP:

$$XP = \sqrt{\frac{A_{mp12}}{A_{mp21}}} \quad \text{Équation 2-57}$$

Ce coefficient XP est donc connu par mesure.

Nous allons maintenant déterminer théoriquement l'expression du coefficient XP en fonction de l'impédance caractéristique des plots,  $Z_p$ .

Dans l'Équation 2-55,  $T_{L1}$ ,  $T_{L2}$  ainsi que le produit  $[T_{m2}] \cdot [T_{m1}]^{-1}$  sont connus, il faut donc exprimer la matrice  $[T_A]$  en fonction de  $Z_p$ .  $[T_A]$  représente la matrice  $[T]$  correspondant à la matrice ABCD, notée  $[A_{AP}]$ , d'un plot de contact décomposé comme montré en Figure 2-27.

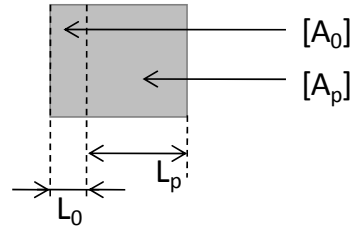


Figure 2-27 : Structure d'un plot de contact reliant une ligne de transmission sous test.

On peut donc écrire :

$$[A_{AP}] = [A_{r1}] \cdot [A_0] \cdot [A_p] \cdot [A_{r2}] \quad \text{Équation 2-58}$$

Avec  $[A_{r1}]$  et  $[A_{r2}]$  les matrices de normalisation, données par les Équation 2-59. Ces matrices permettent de normaliser en entrée à  $Z_0 = 50\Omega$ , impédance de référence due au calibrage OSTL et en sortie à  $Z_p$ , impédance caractéristique des plots de contact.

$$[A_{r1}] = \begin{bmatrix} \frac{1}{\sqrt{Z_0}} & 0 \\ 0 & \sqrt{Z_0} \end{bmatrix}; \quad [A_{r2}] = \begin{bmatrix} \sqrt{Z_p} & 0 \\ 0 & \frac{1}{\sqrt{Z_p}} \end{bmatrix} \quad \text{Équation 2-59}$$

La partie à gauche du plot peut être symbolisée par un circuit ouvert en entrée ramenée par l'impédance caractéristique du plot,  $Z_p$  à la distance  $L_0$ . La matrice ABCD de cette zone est :

$$[A_0] = \begin{bmatrix} \frac{1}{\tanh(\gamma_p \cdot L_0)} & 0 \\ \frac{\tanh(\gamma_p \cdot L_0)}{Z_p} & 1 \end{bmatrix} \quad \text{Équation 2-60}$$

La partie droite du plot, qui sera connectée à la ligne sous test, est considérée comme une ligne de transmission, de longueur  $L_p$ , d'impédance caractéristique  $Z_p$  et d'exposant de propagation  $\gamma_p$ . Sa matrice ABCD est alors :

$$[A_p] = \begin{bmatrix} \cosh(\gamma_p \cdot L_p) & Z_p \cdot \sinh(\gamma_p \cdot L_p) \\ \frac{\sinh(\gamma_p \cdot L_p)}{Z_p} & \cosh(\gamma_p \cdot L_p) \end{bmatrix} \quad \text{Équation 2-61}$$

Grâce aux Équation 2-58 à Équation 2-61 on obtient  $[A_{AP}]$  en fonction de  $[Z_P]$ . En transformant cette matrice ABCD en matrice  $[T]$ , nous obtenons l'expression de la matrice  $[T_A]$  recherchée.

Nous rappelons que nous cherchons à obtenir la forme théorique suivante :  $[T_A] \cdot [T_{L2}] \cdot [T_{L1}]^{-1}$ .  $[T_A]^{-1}$  dans le but d'effectuer une comparaison avec les valeurs issues des mesures.

$[T_{L2}]$ ,  $[T_{L1}]^{-1}$  et  $[T_A]$  sont connus, dans la mesure où  $[T_A]$  est inversible (déterminant non nul), on en déduit que  $[T_A]^{-1}$  est l'inverse de la matrice  $[T_A]$ .

On note,  $[T_{thp}]$  la matrice  $[T]$  théorique globale :

$$[T_{Thp}] = [T_A] \cdot [T_{L2}] \cdot [T_{L1}]^{-1} \cdot [T_A]^{-1} \quad \text{Équation 2-62}$$

En transformant  $[T_{Thp}]$ , en matrice ABCD, noté  $[A_{Thp}]$ , on peut définir le coefficient XP théorique en fonction de  $Z_P$  :

$$XP = \sqrt{\frac{A_{thp12}}{A_{thp21}}} \quad \text{Équation 2-63}$$

A l'aide des Équation 2-57 et Équation 2-63 on obtient l'impédance caractéristique des plots de contacts :

$$Z_P = XP \cdot Z_0 \cdot \sqrt{1 - \tanh(\gamma_p \cdot L_0)^2} \quad \text{Équation 2-64}$$

Pour déterminer l'impédance caractéristique de la ligne de transmission sous test,  $Z_c$ , le principe est le même. Pour cela il faut déterminer la matrice  $[T_A]$  théorique, de la zone A, de la CDE présenté Figure 2-26-b. On compare le coefficient, noté XL maintenant, issu des mesures faites sur les dispositifs de la famille ligne à celui tiré de la théorie (Figure 2-28).

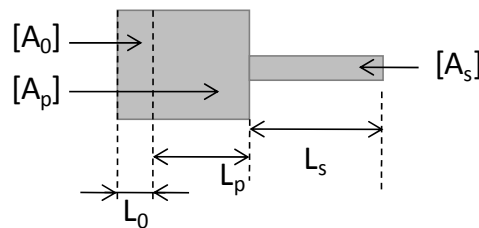


Figure 2-28 : Structure d'un plot de contact et d'un morceau de ligne d'accès

On effectue exactement le même processus que précédemment avec les nouvelles mesures. Ce qui nous donne le coefficient XL, le rapport des deux termes extra diagonaux.

$$XL = \sqrt{\frac{A_{mL12}}{A_{mL21}}} \quad \text{Équation 2-65}$$

Ce coefficient étant connu, il nous reste à définir théoriquement le coefficient XL en fonction de  $Z_c$ .

Nous avons vu, que le système à définir est de la forme  $[T_{LA}].[T_{L2}].[T_{L1}]^{-1}.[T_{LA}]^{-1}$ . Les matrices  $[T_{L2}]$  et  $[T_{L1}]^{-1}$  sont connus issus des mesures, il rest à déterminer la matrice  $[T_{LA}]$ , du dispositif Figure 2-28, en fonction de  $Z_c$ .

Nous notons  $[A_{aL}]$  la matrice ABCD correspondant à la matrice de transfère  $[T_{LA}]$  du dispositif Figure 2-28.

Pour obtenir la matrice ABCD de la structure présenté Figure 2-28, zone A de la Figure 2-26-b, il faut cascader les matrices  $[A_o]$ ,  $[A_p]$  et  $[A_s]$  puis les normaliser en entrée à  $Z_o = 50\Omega$ , impédance de référence dû au calibrage OSTL et en sortie à  $Z_c$ , impédance caractéristique de la ligne, comme donnée en Équation 2-66.

On nommera  $[A_{rL1}]$  et  $[A_{rL2}]$ , respectivement les matrices de référence d'entrée et de sortie.

$$[A_{aL}] = [A_{rL1}].[A_o].[A_p].[A_s].[A_{rL2}] \quad \text{Équation 2-66}$$

Les matrices  $[A_o]$  et  $[A_p]$  sont les mêmes que précédemment, il nous reste à définir la matrice  $[A_s]$  :

$$A_s = \begin{bmatrix} \cosh(\gamma_s.L_s) & Z_c \cdot \sinh(\gamma_s.L_s) \\ \frac{\sinh(\gamma_s.L_s)}{Z_c} & \cosh(\gamma_s.L_s) \end{bmatrix} \quad \text{Équation 2-67}$$

En passant la matrice ABCD  $[A_{aL}]$  en matrice T on obtient la matrice recherché  $[T_{LA}]$ . Cette matrice étant inversible, on obtient la matrice globale théorique  $[T_{thL}]$  (Équation 2-68).

$$[T_{thL}] = [T_{LA}].[T_{L2}].[T_{L1}]^{-1}.[T_{LA}]^{-1} \quad \text{Équation 2-68}$$

Le passage de cette matrice T en matrice ABCD nous permet d'obtenir la matrice théorique  $[A_{thL}]$  en fonction de  $Z_c$  et donc de déterminer le coefficient XL théorique (Équation 2-69).

$$XL = \sqrt{\frac{A_{thL12}}{A_{thL21}}} \quad \text{Équation 2-69}$$

A l'aide des Équation 2-65 et Équation 2-69 on obtient l'impédance caractéristique de la ligne de transmission sous test:



$$Z_c = Z_p \cdot \sqrt{\frac{Num}{Deno}} \quad \text{Équation 2-70}$$

Avec

$$Num = Z_p^2 \cdot \sinh(\gamma_p \cdot L_p)^2 + Z_0^2 \cdot XL^2 (\cosh(\gamma_p \cdot L_p) + \sinh(\gamma_p \cdot L_p) \cdot \tanh(\gamma_p \cdot L_0))^2$$

$$Deno = Z_0^2 \cdot XL^2 (\cosh(\gamma_p \cdot L_p) + \sinh(\gamma_p \cdot L_p) \cdot \tanh(\gamma_p \cdot L_0))^2 \cdot Z_p^2 \cdot \cosh(\gamma_p \cdot L_p)^2$$

### II.3.4.c Domaine de validité

Lors de la détermination de l'impédance caractéristique de l'interconnexion, on suppose que la transition entre les plots de contact et la zone centrale sous test est modélisée par un simple saut d'impédance, conduisant à un coefficient de réflexion. Or il est connu qu'un tel changement d'impédance ne peut se représenter que par ce simple saut d'impédance, mais par une matrice  $S_{ij}$  différente de celle due au saut d'impédance, cela est d'autant plus vrai que le changement de topologie est marqué. Afin de valider les extractions d'impédance caractéristique, nous avons mis en place une procédure permettant de mesurer la matrice  $S_{ij}$  de la transition entre les plots de contacts et l'interconnexion. La matrice  $S_{ij}$  de la transition est obtenue expérimentalement à partir d'un double calibrage LRL réalisée au niveau des plans P et P' (Figure 2-29) sur des structures spécifiques.

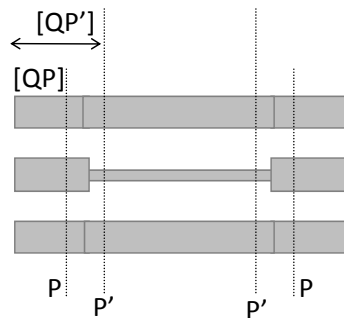


Figure 2-29 : Cellule de test sur laquelle est réalisée un double calibrage LRL, l'un au niveau des plans P, l'autre au niveau des plans P'

A partir des matrices  $[S_{ij}]$  des quadripôles d'erreurs  $[Q_P]$  et  $[Q_{P'}]$  (respectivement obtenus par deux calibrages) que l'on transforme en matrice  $[T]$  notées  $[T_{QP}]$  et  $[T_{QP'}]$ , il est possible d'obtenir la matrice  $[T_t]$  de la transition plot / ligne :

$$[T_{QP'}] = [T_{QP}] \cdot [T_t] \quad \text{Équation 2-71}$$

Une fois la matrice  $[T_t]$  connue, on en déduit la matrice  $[S_t]$  que l'on compare avec celle obtenue avec les mesures de  $Z_{plot}$  et  $Z_c$  d'après la relation suivante :

$$[S_t] = \begin{bmatrix} |\Gamma| & |\sqrt{1-\Gamma^2}| \\ |\sqrt{1-\Gamma^2}| & |\Gamma| \end{bmatrix} \quad \text{Équation 2-72}$$

Si les matrices  $[S_t]$  et  $[S_r]$  sont identiques, la procédure d'extraction mise en place est valide. Dans le cas contraire, cela signifie que la transition entre les plots de contacts et l'interconnexion ne se comporte pas comme un simple saut d'impédance et l'extraction de  $Z_c$  est moins précise due à la mauvaise modélisation de la transition.

La démonstration de la validité de cette hypothèse est donnée par [Roullard-2, 2011].

#### II.3.4.d Conclusion

Cette méthode de de-embedding présente les mêmes avantages et inconvénients que la précédente, à savoir une rigueur théorique mais une lourdeur de mise en œuvre. Ces deux méthodes, proches dans les principes utilisés, diffèrent par la formulation théorique utilisée.

### II.3.5 Conclusion générale sur les méthodes de de-embedding présentées

Nous avons détaillé quatre techniques d'épluchage spécifiques à la caractérisation de dispositifs sous test intégrés et enfouis dans un environnement microélectronique. Ces méthodes nécessitent chacune de réaliser des cellules de de-embedding spécifiques mais qui diffèrent en nombre et par leur topologie. D'autre part les hypothèses nécessaires à la validité de ces méthodes d'épluchage sont de natures différentes, certaines hypothèses portent sur les modèles électriques utilisés, d'autres sur la topologie des structures de test. C'est pour ces raisons pratiques que nous les avons étudiées, programmées, voire améliorées et utilisées. En effet lors de nos campagnes expérimentales nous ne pouvions pas toujours obtenir les surfaces de silicium voulues, par conséquent il fallait minimiser le nombre de dispositifs de de-embedding sur les masques, dans d'autres cas nous ne pouvions pas, par exemple, réaliser autant de niveaux de métallisation que souhaité et par conséquent nous avons des contraintes de design fortes et peu compatibles avec celles nécessaires au design des structures de de-embedding envisagées.

Néanmoins nous allons montrer que lorsque nous pouvons les mettre en œuvre simultanément elles conduisent à des résultats comparables lors de la caractérisation des dispositifs d'interconnexions qui font l'objet de nos investigations.

### III Caractérisation des briques élémentaires d'interconnexions dans un empilement 3D de puces

Cette partie regroupe les résultats de caractérisation et de modélisation des quatre briques élémentaires d'interconnexions que l'on rencontre lors de l'empilement 3D de puces :

- Une interconnexion de surface de type RDL,
- une interconnexion enfouie de type BEOL,
- un via traversant le silicium TSV,
- un pilier de cuivre Cu-pillar

Ainsi que les résultats concernant une inductance 2D intégrée afin d'évaluer les modifications de ces caractéristiques lors de l'empilement.

En comparant les résultats extraits expérimentalement à ceux obtenus par simulation nous pourrions valider les simulateurs électromagnétiques utilisés. Les deux outils de simulation élaborés par ANSYS qui ont été utilisés sont : Maxwell 2D et HFSS. Ces deux logiciels fonctionnent sur une base de calcul par éléments finis. Le logiciel Maxwell 2D, qui présente l'avantage d'avoir un temps de calcul relativement rapide, est utilisé pour la simulation d'interconnexions uniformes dans leur direction de propagation. Cependant pour toutes simulations effectuées dans un environnement plus complexe, comprenant un assemblage 3D ou une chaîne d'interconnexion, il est nécessaire d'utiliser un simulateur de type HFSS plus complexe et nécessitant des temps de calcul beaucoup plus long que Maxwell 2D.

#### III.1 Interconnexions de type RDL (Redistribution Layer).

Dans un premier temps l'impédance caractéristique d'une interconnexion typique de celle rencontrée pour les lignes de redistribution (RDL) est extraite par le biais de trois méthodes de de-embedding présentées dans le paragraphe précédent (Méthode de DF Williams, de TM Winkel et celle dite d'épluchage par bloc mise en place par S. De Rivaz). Les résultats obtenus sont confrontés à celui obtenu avec le simulateur Maxwell 2D. Cette étude nous permet de confronter nos procédures de caractérisation et de modélisation. Dans un second temps nous caractérisons une interconnexion de RDL, dans un environnement d'intégration 3D, avec la technique de de-embedding de D.F. Williams (nous ne disposons pas de toutes les cellules de de-embedding pour appliquer les autres techniques).

##### III.1.1 Description du dispositif de test associé à une interconnexion de type RDL

Ce véhicule de test présente un empilement technologique simple. Il est composé d'un substrat de verre BK7 et les dispositifs déposés sur ce substrat sont en aluminium (Figure 2-30). Les lignes de

transmission ont une épaisseur de 220 nm ainsi que plusieurs largeurs ( $d$ ). Deux largeurs de 20  $\mu\text{m}$  et 40  $\mu\text{m}$  seront étudiées. Les masses ont une épaisseur de 220 nm et une largeur de 80  $\mu\text{m}$ .

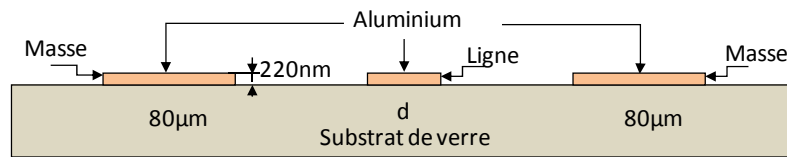


Figure 2-30 : Empilement technologique du dispositif étudié.

La Figure 2-31 montre un « Mapping » du véhicule de test, où une grande quantité de cellules de de-embedding permettant l'application de nos méthodes d'épluchage est disponible.

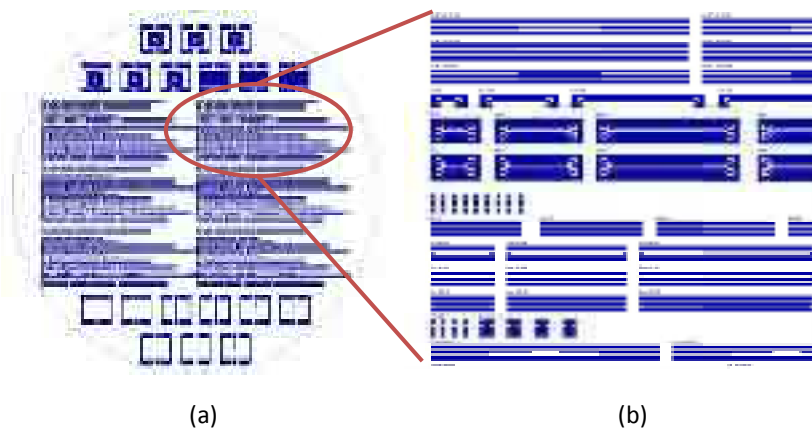


Figure 2-31 : « Mapping » du véhicule de test (a) vue d'ensemble (b) zoom sur les lignes de transmission.

### III.1.2 Résultats d'extraction et de modélisation

La partie réelle et la partie imaginaire de l'impédance caractéristique des lignes de transmission sous test sont présentées en fonction de la fréquence dans la Figure 2-32.

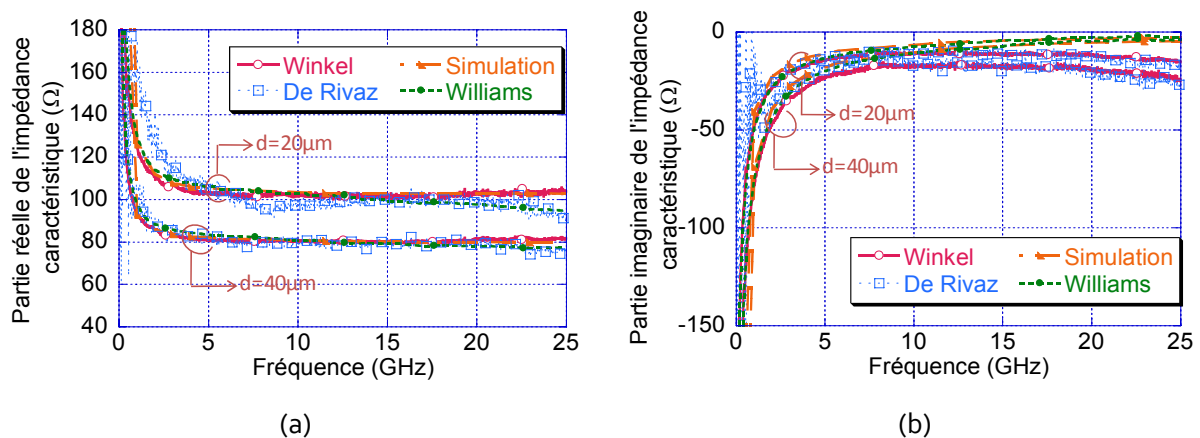


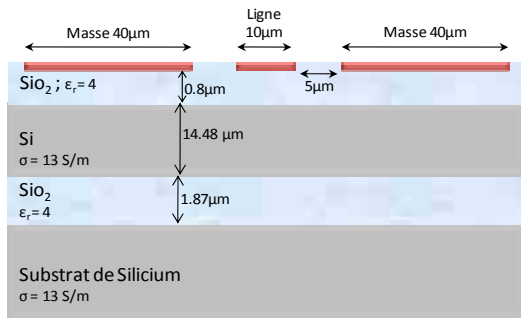
Figure 2-32 : Impédance caractéristique des lignes de transmission de largeur 20 et 40  $\mu\text{m}$ .  
(a) Partie réelle (b) Partie imaginaire.

On remarque que toutes les méthodes d'extractions donnent des résultats très proches de la simulation sur une large bande de fréquence. Toutefois la méthode de S. De-Rivaz conduit à des résultats un peu plus bruités en basse fréquence que les autres, ceci étant dû aux multiples mesures à effectuer. La très bonne concordance entre les résultats obtenus avec les trois méthodes de De-Embedding et avec le simulateur Maxwell 2D montre la pertinence de nos outils de caractérisation et de modélisation.

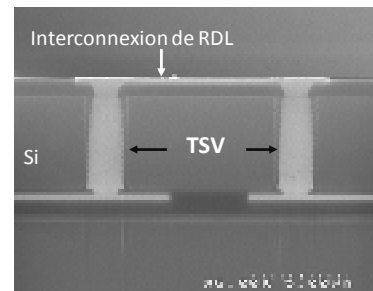
### III.1.3 Interconnexion de RDL dans un environnement réel d'intégration 3D

#### III.1.3.a Description du dispositif sous test et de sa cellule de test

Les mesures ont été effectuées sur une cellule de test comprenant le dispositif sous test des lignes d'accès ainsi que des plots de contact. L'empilement technologique est composé d'un substrat de silicium (Si), de deux couches de dioxyde de silicium ( $\text{SiO}_2$ ) et d'une seconde couche de silicium (Figure 2-33-a). Les conducteurs, d'épaisseur 300 nm, sont situés en haut de la seconde couche de  $\text{SiO}_2$ . Le ruban signal a une largeur de  $10\mu\text{m}$  et les rubans de masses ont des largeurs de  $40\mu\text{m}$ . La méthode de de-embedding utilisée est celle de D.F.Williams.



(a)



(b)

Figure 2-33 : Interconnexion de RDL dans un environnement d'intégration 3D.  
(a) Coupe schématisée. (b) Coupe MEB de l'empilement technologique.

#### III.1.3.b Résultats

La Figure 2-34 nous montre les paramètres primaires associés au modèle électrique classique d'une ligne de transmission (Figure 2-35). Les résultats expérimentaux de ces deux structures sont confrontés à des résultats issus de simulation électromagnétique.

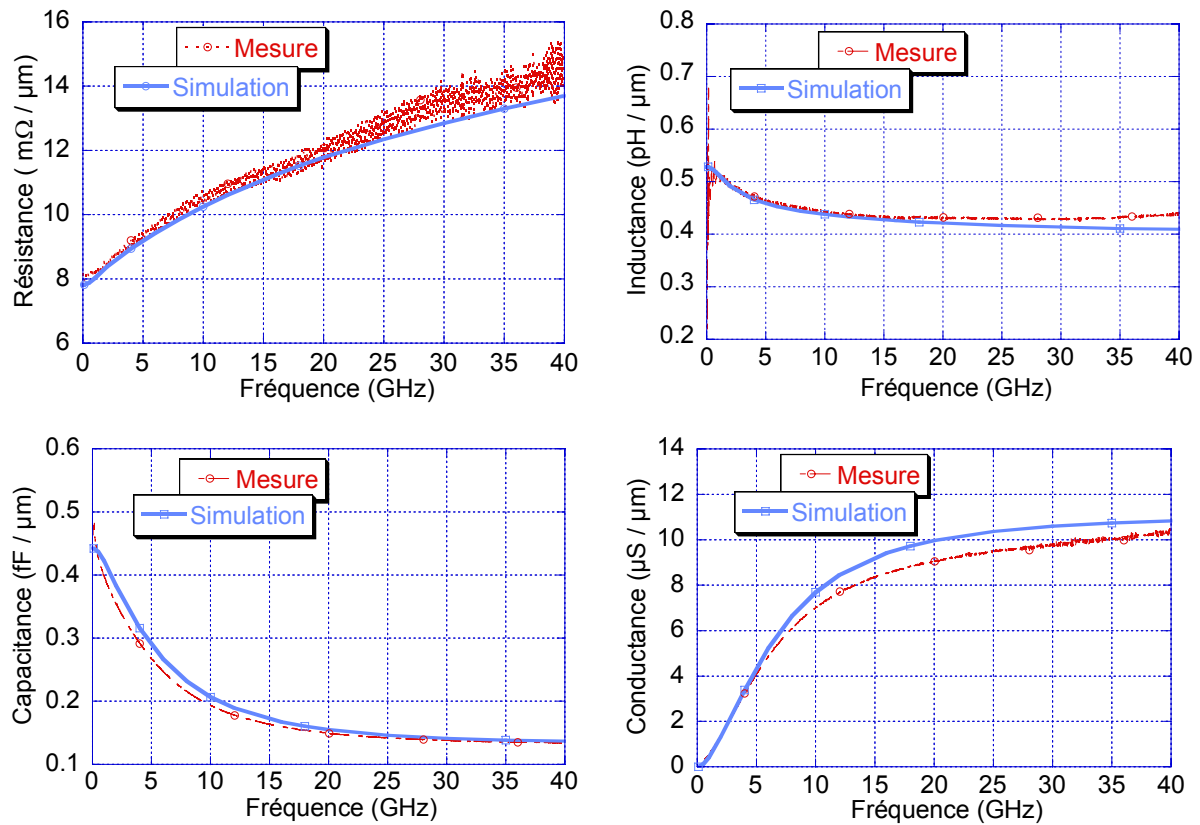


Figure 2-34 : Paramètres R, L, C et G d'une interconnexion de RDL

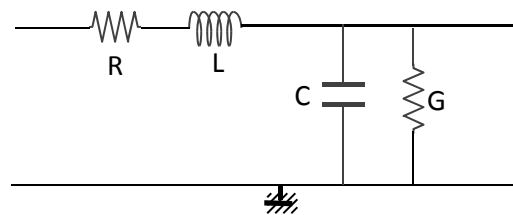


Figure 2-35 : Modèle électrique classique d'une ligne de transmission.

L'augmentation du paramètre R est majoritairement due à l'effet de peau dans les conducteurs.

Le paramètre L diminue légèrement en fonction de la fréquence. A haute fréquence les champs électriques et magnétiques sont de plus en plus confinés sur la périphérie des conducteurs ce qui tend à diminuer l'inductance interne des conducteurs.

Les paramètres C et G ne sont pas constants sur toute la bande fréquentielle à cause de la pénétration des champs électromagnétiques dans le substrat de silicium. C'est pourquoi nous proposons Figure 2-36, un modèle électrique équivalent de la partie parallèle de la ligne de transmission. Le paramètre C est composé de deux contributions : la capacité d'oxyde et la capacité de silicium. Le paramètre G représente les pertes diélectriques dans le silicium.

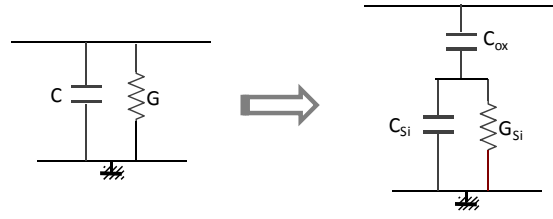


Figure 2-36 : Modèle électrique de la partie parallèle d'une interconnexion ; Côte gauche le modèle classique et côté droit le modèle équivalent.

Les paramètres primaires sont extraits à partir des graphiques Figure 2-34, R et L sont définis à partir des valeurs statiques tandis que  $C_{ox}$ ,  $C_{si}$  et  $G_{si}$  sont déterminés à l'aide du modèle équivalent Figure 2-36. En effet, en exprimant le modèle classique et le modèle équivalent sous la forme d'admittance, respectivement  $Y_{classique}$  (Équation 2-73 et Figure 2-35) et  $Y_{équivalent}$  (Équation 2-74) et en procédant par identification (Équation 2-75 et Équation 2-76) on obtient les paramètres  $C_{ox}$ ,  $C_{si}$  et  $G_{si}$  en fonction de C et G (Équation 2-77 à Équation 2-79).

$$Y_{classique} = G + jC\omega \quad \text{Équation 2-73}$$

$$Y_{équivalent} = \frac{\omega^2 \cdot G_{si}^2 \cdot C_{ox}^2}{G_{si}^2 + \omega^2 (C_{ox} + C_{si})^2} + j\omega \frac{C_{ox}G_{si} + \omega^2 C_{ox}C_{si}(C_{ox} + C_{si})}{G_{si}^2 + \omega^2 (C_{ox} + C_{si})^2} \quad \text{Équation 2-74}$$

$$G = \frac{\omega^2 \cdot G_{si}^2 \cdot C_{ox}^2}{G_{si}^2 + \omega^2 (C_{ox} + C_{si})^2} \quad \text{Équation 2-75}$$

$$C = \frac{C_{ox}G_{si} + \omega^2 C_{ox}C_{si}(C_{ox} + C_{si})}{G_{si}^2 + \omega^2 (C_{ox} + C_{si})^2} \quad \text{Équation 2-76}$$

$$\text{A basse fréquence } (\omega \rightarrow 0); \quad C_{ox} = C \quad \text{Équation 2-77}$$

$$\text{A haute fréquence } (\omega \rightarrow \infty); \quad C_{si} = \frac{C_{ox}C}{C_{ox}-C} \quad \text{Équation 2-78}$$

$$\text{A haute fréquence } (\omega \rightarrow \infty); \quad G_{si} = \frac{G(C_{ox}+C_{si})^2}{C_{ox}^2} \quad \text{Équation 2-79}$$

Le tableau ci-dessous synthétise les valeurs des paramètres primaires des deux structures. On note respectivement R et L les valeurs de résistance et d'inductance en statique,  $C_{ox}$  la valeur de la capacité d'oxyde,  $C_{si}$  la valeur de la capacité dans le silicium et G les pertes causées par le substrat de silicium.

Tableau 3-1 : Valeurs des paramètres primaires des interconnexions de RDL.

$R$ (mΩ/μm)	$L$ (pH/μm)	$C_{ox}$ (fF/μm)	$C_{si}$ (fF/μm)	$G_{si}$ (μS/μm)
8	0.51	0.45	0.2	19.4

### III.2 Interconnexion du BEOL enfouie dans un environnement 3D

Dans cette partie nous nous intéressons à l'extraction du modèle RLCG d'une interconnexion enfouie dans un BEOL silicium après empilement 3D. Les résultats obtenus sont comparés à ceux issus d'une simulation full-wave avec HFSS. Les dispositifs de de-embedding disponibles avec cette interconnexion nous permettent d'appliquer les méthodes d'épluchage suivantes :

- Méthode de T. Bertaud dérivée de celle développée par E.P. Van Damme
- Méthode développée par D.F. Williams.
- Méthode DUMMC développée par L. Fourneaud

Les autres méthodes, nécessitant des CDE spécifiques mais non prévus sur le véhicule de test utilisé, ne peuvent être mises en œuvre ici.

#### III.2.1 Description du dispositif sous test

Les mesures ont été effectuées sur une famille d'interconnexions du BEOL silicium réalisées avec plusieurs longueurs. Ces interconnexions sous test sont enfouies dans une cellule de test possédant des plots de contact, des lignes d'accès et des TSV. L'application des différentes méthodes d'épluchage mises en œuvre nous permet d'extraire les valeurs de éléments R, L, C et G du modèle équivalent à une ligne de transmission. Le caractère localisé des TSV ainsi que la présence de CDE adéquate permettent d'utiliser la méthode de de-embedding de D.F. Williams.

L'empilement technologique est composé d'un substrat de silicium (conductivité 13 S/m), d'une couche de dioxyde de silicium (épaisseur 1.87  $\mu\text{m}$ ), d'une seconde couche de silicium (épaisseur 14.48  $\mu\text{m}$  et conductivité 13 S/m), et d'une seconde couche de dioxyde de silicium (épaisseur 1.2  $\mu\text{m}$ ) comprenant les masses en cuivre (épaisseur 192 nm). L'interconnexion (épaisseur 300 nm) est située dans la couche de dioxyde de silicium situé entre les deux substrats et à une distance de 135 nm du silicium supérieur (Figure 2-37).

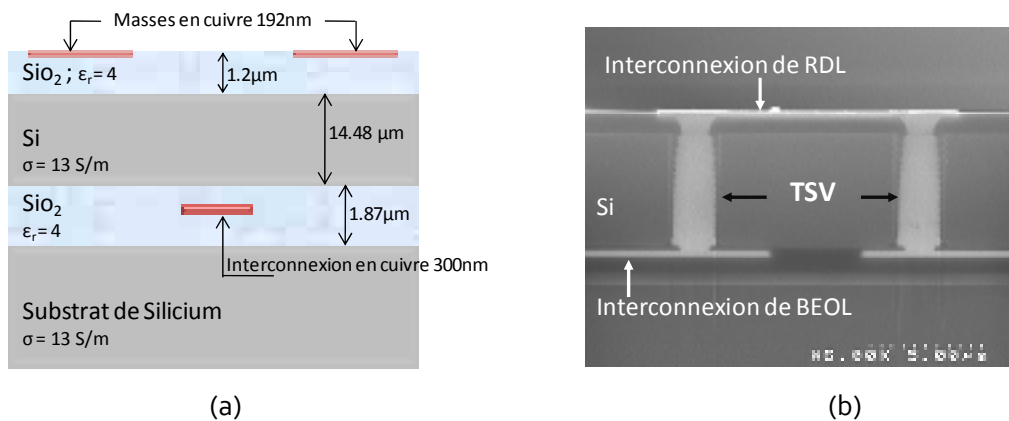


Figure 2-37 : Interconnexion enfouie dans un BEOL silicium en intégration 3D.  
(a) Coupe schématique. (b) Coupe MEB de l'empilement technologique.



### III.2.2 Résultats d'extraction et de modélisation

En appliquant les méthodes de de-embedding explicitées précédemment on obtient les parties réelle et imaginaire de l'impédance caractéristique des lignes de transmission sous test Figure 2-38.

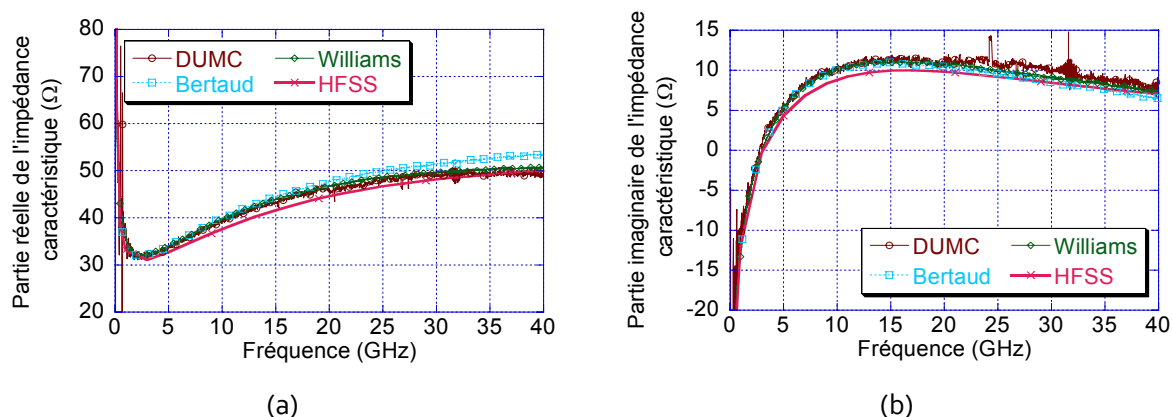


Figure 2-38 : Impédance caractéristique d'interconnexion enfouie dans un BEOL silicium en intégration 3D. (a) Partie réelle (b) Partie imaginaire.

On remarque que toutes les méthodes d'extractions donnent des résultats nets et très proches de la simulation sur l'ensemble de la large bande de fréquence. De plus grâce aux Équation 2-19 et Équation 2-20, on peut remonter aux paramètres linéiques R, L, C et G de la ligne de transmission sous test, présentés en Figure 2-39.

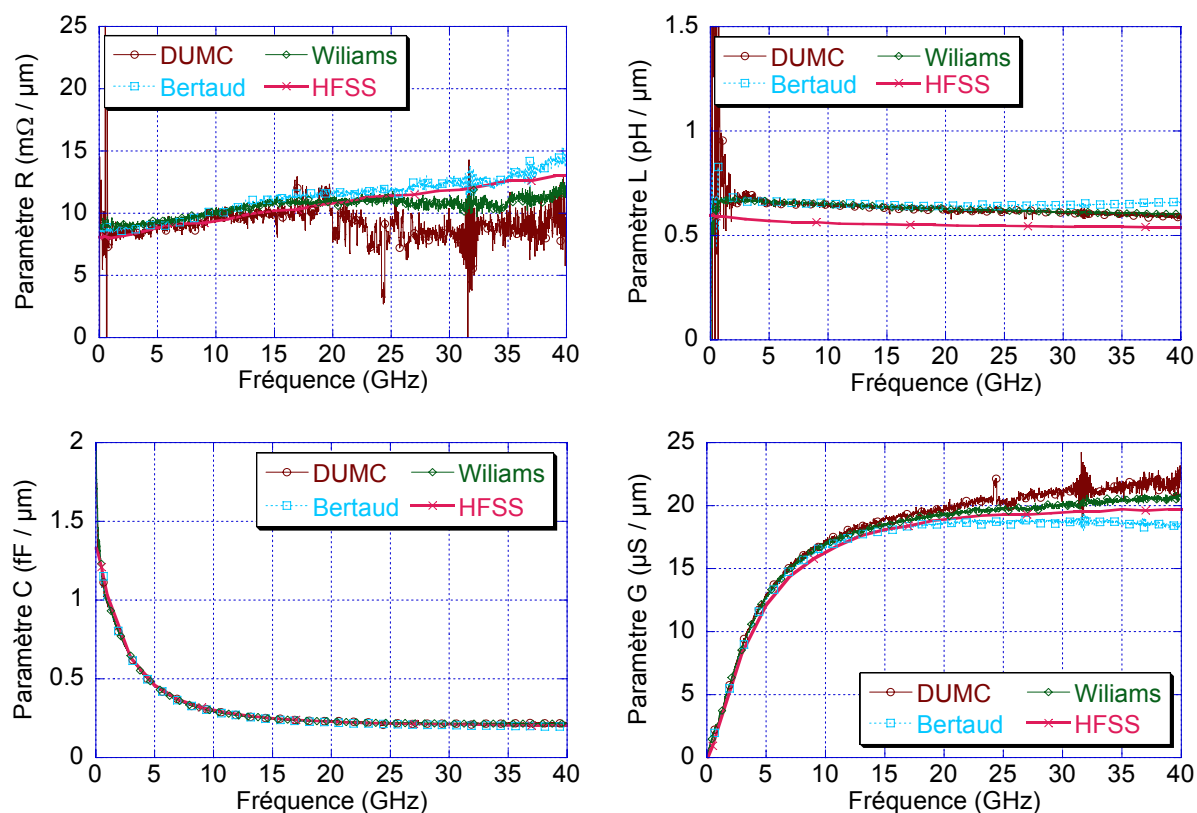


Figure 2-39 : Paramètres R, L, C et G de la ligne de transmission

Les évolutions fréquentielles de R, L, C et G sont typiques de ceux attendus pour d'une ligne de transmission avec pertes métalliques et diélectriques.

Le paramètre R (Figure 2-39), augmente en fonction de la fréquence, ceci est dû à l'effet de peau dans les conducteurs caractérisé par son épaisseur de peau  $\delta$  (Équation 2-80). Ce phénomène provoque la décroissance de la densité de courant à mesure que l'on s'éloigne de la périphérie du conducteur. Il est habituel de dire que le courant circule principalement dans l'épaisseur  $\delta$  à la périphérie des métallisations.

$$\delta = \sqrt{\frac{2}{\omega \cdot \mu \cdot \sigma}} \quad \text{Équation 2-80}$$

Avec :  $\omega$ , la pulsation (rad/s).  $\mu$ , la perméabilité magnétique (H/m).  $\rho$ , la résistivité ( $\Omega \cdot m$ ) et  $\sigma$  la conductivité (S/m).

La valeur de la résistance statique moyenne est égale à 8300  $\Omega/m$ , ce qui se vérifie facilement par l'Équation 2-81 :

$$R_{dc} = R_{\text{ligne}} + \frac{R_{\text{masse}}}{2} = \rho \frac{1}{w_{\text{ligne}} \cdot t} + \frac{\rho}{2} \frac{1}{(w_{\text{masse}} \cdot t)} = 8029.5 \Omega/m \quad \text{Équation 2-81}$$

L'écart entre la valeur théorique et la valeur expérimentale est faible, de l'ordre de 3,6% et il peut s'expliquer par l'incertitude que nous avons quant à la géométrie exacte de la structure de test et de conductivité des matériaux. Dans notre cas nous avons estimé la conductivité du cuivre à  $4.8 \cdot 10^7$  S/m, la largeur de la ligne à 10  $\mu m$ , son épaisseur à 300 nm, la largeur des masses à 50  $\mu m$  et leurs épaisseurs à 192 nm. À partir de ces données, nous constatons une très bonne concordance entre la simulation et les mesures.

L'inductance L diminue légèrement en fonction de la fréquence. En effet à haute fréquence et à cause de l'effet de peau les champs électriques et magnétiques sont de plus en plus confinés sur la périphérie des conducteurs ce qui tend à diminuer l'inductance interne des conducteurs. On note également que la simulation sous-estime ce paramètre d'environ 9%.

La capacitance C et la conductance G ne sont pas constants à cause de la pénétration des champs électromagnétiques dans le substrat de silicium. La capacitance C est composée des contributions des capacités d'oxyde et de silicium. La conductance G représente les pertes diélectriques dans le silicium. Pour ces deux grandeurs on observe également une très bonne concordance entre leurs valeurs mesurées et modélisées.

Les interconnexions du BEOL présentent des topologies plus agressives que celles de RDL et par conséquent des niveaux de pertes plus élevés. En contre partie elles sont généralement plus

courtes. Par conséquent il est difficile à ce stade de l'étude de conclure quant à leur impact sur les performances de transmission d'un signal en termes de retard, de temps de montée ou niveau de signal transmis.

Nous nous intéressons maintenant aux interconnexions verticales présentes en intégration 3D.

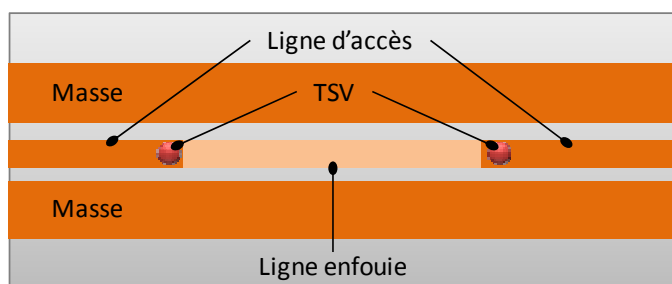
### III.3 Through Silicon Via (TSV) haute densité

Après avoir décrit les dispositifs de test mesurés, un protocole permettant de déterminer les valeurs des éléments du modèle électrique d'un TSV est exposé. Enfin les résultats obtenus et issus des mesures et des simulations électromagnétiques seront comparés.

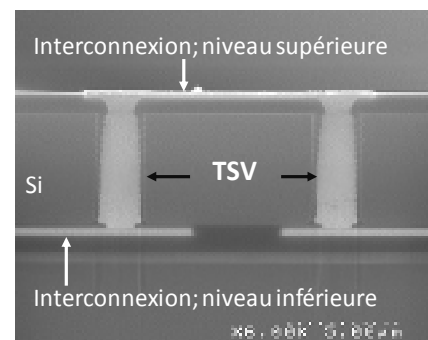
#### III.3.1 Description de la cellule de test mesurée

La cellule de test incluant des TSV, enfouis en son sein, est constituée de :

- Plots et lignes d'accès : Ils font le lien entre les sondes de mesure HF et les TSV. Ces lignes d'accès sont les RDL que nous venons de caractériser.
- TSV à haute densité : son diamètre est  $3\ \mu\text{m}$  et sa hauteur de  $15\ \mu\text{m}$ . Ses caractéristiques technologiques sont les suivantes :
  - Isolation diélectrique en  $\text{SiO}_2$ , épaisseur  $200\ \text{nm}$
  - Barrière en  $\text{TIN}$ , épaisseur  $20\ \text{nm}$
  - Métallisation en cuivre, de diamètre  $3\ \mu\text{m}$
- Interconnexion du BEOL enfouie : Elle permet la connexion entre les deux TSV sur la face du BEOL. Nous venons de les caractériser.



(a)



(b)

Figure 2-40 : vue de la cellule de test (a) De dessus. (b) En coupe.

La Figure 2-41 présente une coupe transversale des TSV haute densité à caractériser.

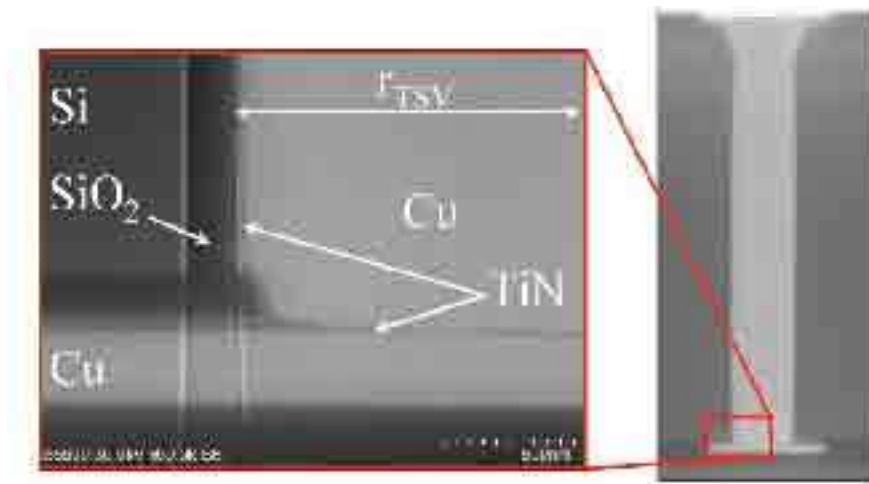


Figure 2-41 : Image d'un TSV haute densité.

On y retrouve les couches successives de SiO<sub>2</sub>, de TiN et de Cuivre.

### III.3.2 Méthode d'épluchage appliquée pour la caractérisation du TSV

Dans un premier temps, nous appliquons la procédure de De-Embedding développée par E.P. Vandamme et modifiée par T. Bertaud pour obtenir la matrice ABCD de l'ensemble constitué par la chaîne TSV+interconnexion du BEOL+TSV, qui constitue notre dispositif sous test enfoui. Nous sommes obligé de passer par cette étape intermédiaire puisque nos accès sur les ports d'entrée et de sortie sont tous les deux situés sur la face supérieure.

La cascade des matrices ABCD des éléments TSV, interconnexion du BEOL et TSV constituant ce DST, Équation 2-82, permet par le biais d'opération matricielle de remonter à la matrice ABCD d'un seul TSV, Équation 2-83.

$$[DST] = [TSV]. [BEOL]. [TSV] \quad \text{Équation 2-82}$$

$$[TSV] = [BEOL]^{-1}. ([BEOL]. [DST])^{1/2} \quad \text{Équation 2-83}$$

[DST] est connu, issu des mesures, on l'obtient grâce à au De-Embedding de type R.P. Vandamme et [Ligne<sub>E</sub>] est la matrice ABCD de la ligne d'interconnexion du BEOL enfouie que nous avons déjà caractérisée.

### III.3.3 Extractions dans les domaines fréquentiel

Les valeurs des éléments du modèle physique proposé par [Cadix, 2009], pour un TSV haute densité sont extraites en fonction de la fréquence.

Dans un premier temps, un modèle électrique générique (Figure 2-42-a) [Pozar, 2005] permet d'associer des paramètres de la matrice [ABCD] du TSV à des admittances complexes, Équations 2-84. Puis un modèle plus physique (Figure 2-42-b) lie ces admittances à des éléments électriques RLCG représentant le comportement physique du TSV, Équation 2-85 à Équation 2-90.

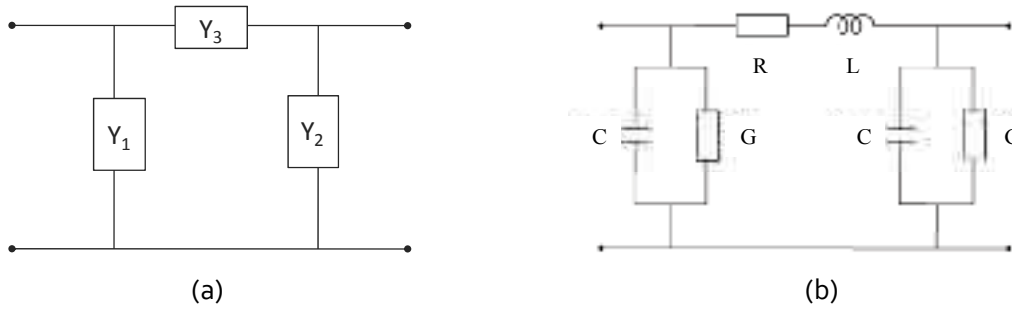


Figure 2-42 : Modèle électrique d'un TSV. (a) Classique. (b) Physique

$$A = 1 + \frac{Y_2}{Y_3}; \quad B = \frac{1}{Y_3}$$

$$C = Y_1 + Y_2 + \frac{Y_1 Y_2}{Y_3}; \quad D = 1 + \frac{Y_1}{Y_3}$$

Équations 2-84

$$R = \text{real}(B)$$

Équation 2-85

$$L = \frac{\text{imag}(B)}{\omega}$$

Équation 2-86

$$G_1 = \text{real}\left(\frac{A-1}{B}\right)$$

Équation 2-87

$$C_1 = \frac{\text{imag}}{\omega}\left(\frac{A-1}{B}\right)$$

Équation 2-88

$$G_2 = \text{real}\left(\frac{D-1}{B}\right)$$

Équation 2-89

$$C_2 = \frac{\text{imag}}{\omega}\left(\frac{D-1}{B}\right)$$

Équation 2-90

La confrontation entre ces paramètres issus des mesures et ceux issus de la simulation électromagnétique est présentée en Figure 2-43.

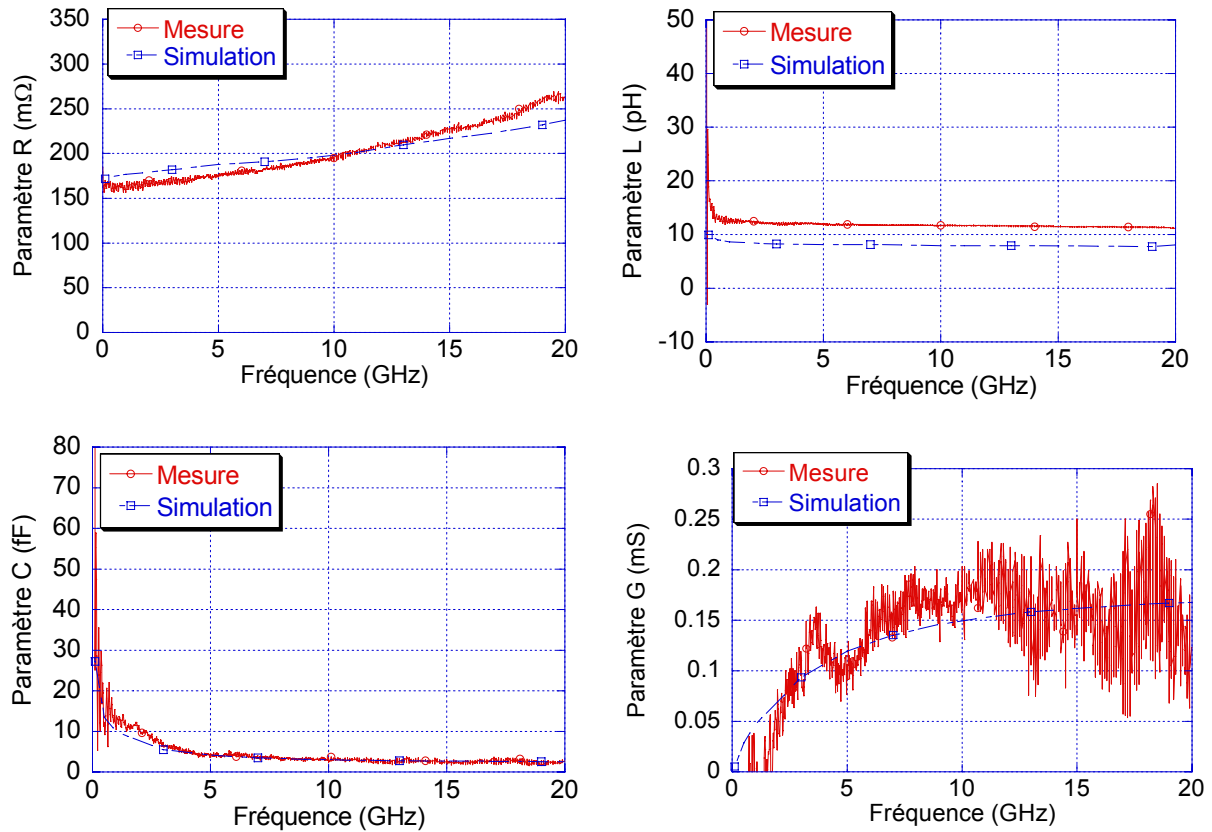


Figure 2-43 : Paramètres primaires d'un TSV haute densité.

Les éléments  $R$ ,  $L$ ,  $C$  et  $G$  traduisent bien les phénomènes physiques attendus pour un TSV haute densité. L'effet de peau fait augmenter la résistance  $R$  et diminuer l'inductance  $L$  à mesure que la fréquence augmente. La capacitance  $C$  et la conductance  $G$  ne sont pas constantes sur toute la bande fréquentielle à cause de la pénétration des champs électromagnétiques dans le substrat de silicium. C'est pourquoi, un modèle électrique équivalent de l'admittance parallèle du TSV est proposé en Figure 2-44. La capacitance  $C$  intègre deux contributions : la capacité liée à la couche d'oxyde et la capacité liée au substrat de silicium. La conductance  $G$  représente les pertes diélectriques dans le silicium. Les valeurs des éléments RLCG du modèle électrique d'un TSV sont définies dans le Tableau 2-2.

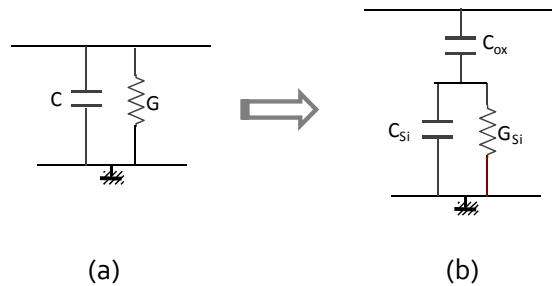


Figure 2-44 : Modèle électrique de la partie parallèle; (a) le modèle classique (b) le modèle équivalent.

Tableau 2-2 : Valeurs caractéristiques d'un TSV haute densité.

	$R \text{ (m}\Omega\text{)}$	$L \text{ (pH)}$	$C_{ox} \text{ (fF)}$	$C_{si} \text{ (fF)}$	$G_{si} \text{ (mS)}$
<b>TSV</b>	165	15	30	2,7	0,2

On observe, dans un premier temps, une très bonne concordance entre la mesure et la simulation sur une large bande de fréquence. Une très bonne extraction des caractéristiques électriques des TSV haute densité, ayant des dimensions très faibles, est ici démontrée.

Remarque : l'effet de la polarisation du substrat a été mis en évidence avec la création d'une capacité de déplétion liée à la valeur de conductivité du silicium [Cadix, 2010]

La même étude, pour des TSV moyenne densité, de diamètre 60  $\mu\text{m}$  et de hauteur 120  $\mu\text{m}$ , a été effectuée par L. Fourneaud, 2011 [Fourneaud-2, 2011].

### III.4 Cu-Pillar

Les dispositifs à caractériser n'étant parvenus au laboratoire que très récemment, les mesures sont toujours en cours.

Nous espérons pouvoir présenter les résultats lors de la soutenance.

### III.5 Inductances planaires intégrées

Les inductances planaires peuvent être considérées comme des interconnexions de topologie particulière. Elles sont présentes dans les réseaux d'interconnexions sur les niveaux supérieurs du BEOL et dans le futur il est envisagé de les intégrer au niveau des lignes de redistribution (RDL), voire dans les interposers. Nous avons eu l'opportunité de disposer de prototype d'inductances planaires réalisées avec ces technologies d'intégration 3D. Nos objectifs ont alors été de caractériser ces inductances intégrées et enfouies dans les puces, en termes de composant passif spécifique, et cela grâce à nos outils d'épluchage puis de mener des études dédiées à l'évaluation de l'impact des substrats de silicium rapportés lors d'une intégration 3D de puces, sur leur facteur de qualité.

Dans cette partie une description des dispositifs mesurés est donc d'abord donnée, suivie de la présentation de la méthodologie d'extraction permettant d'obtenir le facteur de qualité de l'inductance.

### III.5.1 Description des inductances intégrées sous test

L'inductance 2 tours présentée en Figure 2-45-a est intégrée sur le niveau RDL, elle a une longueur de 480  $\mu\text{m}$ , une largeur de 320  $\mu\text{m}$  et l'espacement entre deux spires est de 20  $\mu\text{m}$ . L'inductance présentée sur la Figure 2-45-b est double ; il y a deux inductances similaires, une sur le niveau RDL et une sur le niveau BEOL. Elles ont les mêmes dimensions que l'inductance simple et elles sont connectées par deux TSV de moyenne densité, de diamètre 60  $\mu\text{m}$  au travers du substrat de silicium d'épaisseur 120  $\mu\text{m}$ .

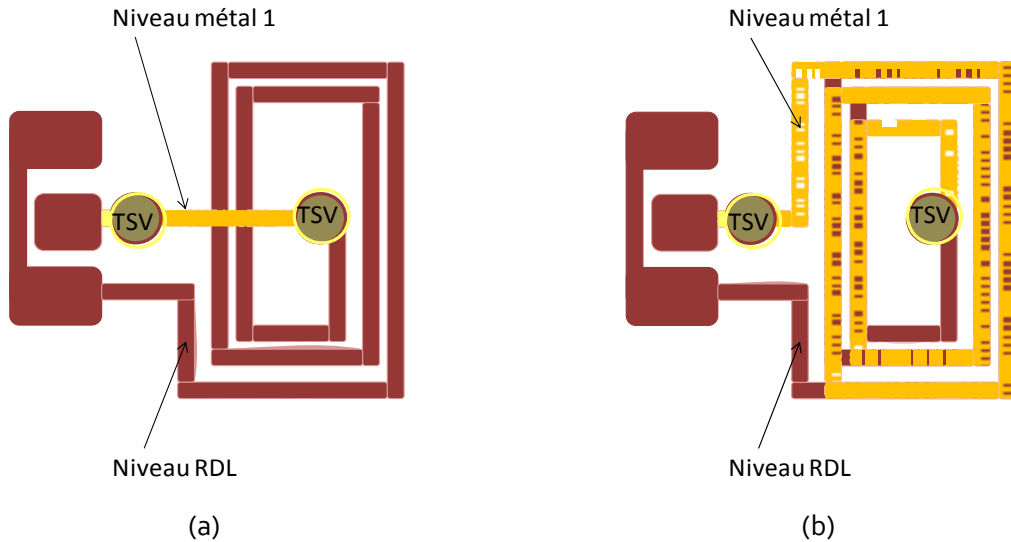


Figure 2-45 : Topologies des inductances intégrées sur la couche RDL. (a) Inductance simple. (b) Inductance double.

### III.5.2 Résultats d'extraction

Après une procédure de de-embedding de type de celle développée par R.P. Van Damme mais simplifiée du fait des uniques mesures en réflexion ( $S_{11}$ ) il devient possible de comparer le facteur de qualité de l'inductance simple à celui d'une inductance double. Le facteur de qualité est le principal critère de qualité des performances des éléments passifs réactifs, qu'ils soient capacitifs, inductifs ou même constitués d'un circuit résonnant.

Dans un premier temps il est nécessaire de supprimer les effets des plots de contact de la cellule de test (CDT). Pour cela on applique le modèle représenté en Figure 2-46, il est constitué de deux admittances en parallèle, la première correspondant aux plots de contact et la seconde à l'inductance.



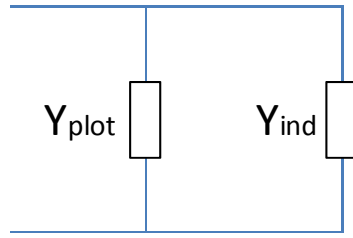


Figure 2-46 : Modèle électrique de l'inductance insérée dans sa cellule de test présentant un plot de contact

L'admittance totale issue de la mesure (Équation 2-91) est la somme des admittances dues aux plots de contact et à l'inductance elle-même.

$$Y_{mesure} = Y_{plot} + Y_{ind} \quad \text{Équation 2-91}$$

$Y_{ind}$  est l'admittance recherchée. Elle s'obtient grâce à l'équation suivante :

$$Y_{ind} = Y_{mesure} - Y_{plot} \quad \text{Équation 2-92}$$

Etant référencé à 50 ohm,  $Y_{mesure}$  et  $Y_{plot}$  sont obtenus par les mesures de  $S_{11}$  sur la cellule de test contenant le plot et l'inductance ( $S_{11_{CDT}}$ ) puis la mesure de  $S_{11}$  de la cellule de de-embedding CDE contenant uniquement le plot ( $S_{11_{plot}}$ ). L'Équation 2-93 et l'Équation 2-94 donnent les relations à utiliser.

$$Y_{mesure} = \frac{1 - S_{11_{CDT}}}{50 (1 + S_{11_{CDT}})} \quad \text{Équation 2-93}$$

$$Y_{plot} = \frac{1 - S_{11_{plot}}}{50 (1 + S_{11_{plot}})} \quad \text{Équation 2-94}$$

Le facteur de qualité est défini comme étant le rapport de l'énergie électromagnétique emmagasinée sur l'énergie dissipée pendant une période de signal, dans l'hypothèse d'une excitation sinusoïdale [Kajfez, 1986]. Pour une inductance ce facteur de qualité est donné par l'Équation 2-95.

$$Q = \frac{-\text{imag}(Y_{11})}{\text{real}(Y_{11})} \quad \text{Équation 2-95}$$

La Figure 2-47 montre que pour un encombrement et une fréquence égale à 2.4 GHz, le facteur de qualité d'une inductance double est deux fois plus élevé que celui d'une inductance simple. La fréquence de coupure haute est aussi largement repoussée.

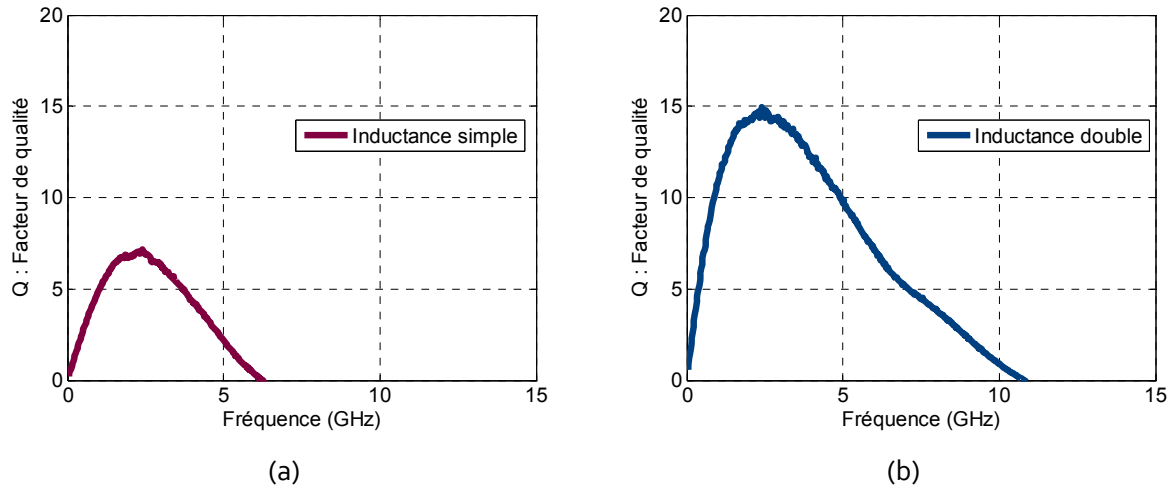


Figure 2-47 : Facteur de qualité. (a) Inductance simple. (b) Inductance double.

La possibilité d'intégrer des inductances sur le niveau RDL donne de réelles perspectives aux designers. En effet avant l'intégration 3D, la superposition des inductances était restreinte aux différents niveaux de métallisation du BEOL, ce qui induisait de forte capacité parasite de couplage entre les spires situées sur ces différents niveaux de métallisation. Avec une architecture de type inductance double, les deux dispositifs inductifs sont largement séparés par le substrat, d'épaisseur 120  $\mu\text{m}$  dans notre cas, ce qui réduit fortement le couplage parasite capacitif. De plus le substrat silicium est de haute résistivité, à 5  $\text{K}\Omega\cdot\text{cm}$ , ce qui tend aussi à limiter les pertes par induction magnétique dans ce dernier.

## IV Conclusion

Dans la première partie de ce chapitre, l'importance des techniques d'épluchage permettant d'extraire les caractéristiques intrinsèques des dispositifs sous test enfouis dans un environnement complexe, a été mise en avant. Nous avons présenté une méthode de de-embedding générique applicable à tous types de dispositifs sous test. Ensuite quatre méthodes de de-embedding bien adaptées à l'extraction de l'impédance caractéristique d'interconnexions enfouie dans un circuit intégré ont été exposées et détaillées. Selon les éléments que nous avons à caractériser (interconnexions de RDL et du BEOL, TSV, Cu-Pillar, inductances) nous avons du utiliser ces différentes techniques car nous ne disposons pas toujours de la place suffisante pour réaliser le nombre de cellules de de-embedding nécessaires à leur mise en œuvre, ni du nombre de niveaux de métallisation nécessaire à leur fabrication. Le Tableau 2-3, rappelle les contraintes en termes de nombre de cellule de de-embedding nécessaires pour ces différentes méthodes d'épluchage.

Tableau 2-3 : Synthèse des méthodes de de-embedding

Auteurs	Nombre de cellules de De-Embedding nécessaires
E.P. Van Damme	3
E.P. Van Damme	2
modifiée par T. Bertaud	
D.F. Williams	3
L. Fourneaud	2
DUMMC	
Par bloc	4
S. De Rivaz	
T.M. Winkel	4
T.M Winkel	4
Modifiée par J. Roullard	

Dans la seconde partie nous avons présenté les résultats de caractérisation obtenus avec ces méthodes de de-embedding sur les différentes briques élémentaires d'interconnexion composant une chaîne d'interconnexion complète dans un empilement 3D de puces. Les interconnexions de RDL, les interconnexions de BEOL, les vias traversant le silicium et les piliers de cuivre ont pu être étudiés et les modèles extraits comparés à ceux obtenus par simulation électromagnétique. Un composant passif réactif a également été caractérisé et a permis de mettre en évidence l'utilité de pouvoir intégrer des inductances sur la couche de redistribution.

Les bons accords obtenus entre résultats expérimentaux et résultats de calcul nous permettent de conclure quant à la fiabilité des outils de caractérisation et de modélisation. Ces derniers pourront être utilisés pour mener des études paramétriques sur chaque brique élémentaire d'interconnexion afin de prédire leur comportement électrique en fonction de leurs dimensions et des caractéristiques des différents matériaux isolant, semi-conducteur et conducteur qui les constituent.



# Chapitre 3

## Modèles et analyses des caractéristiques électriques des briques élémentaires d'interconnexions dans les empilements 3D de circuits intégrés

---

Dans ce chapitre, les comportements électriques de briques élémentaires d'interconnexions pouvant constituer une chaîne complète d'interconnexions dans un empilement 3D de puces sont étudiées dans les domaines fréquentiel et temporel.

La première partie est consacrée aux interconnexions horizontales. Dans un premier temps, les comportements en fréquence ainsi que le retard induit par différentes interconnexions de redistribution sont étudiés. Puis une étude fréquentielle et six règles d'optimisation, au niveau du design, sont déterminées pour les interconnexions de BEOL type 45 nm.

La seconde partie concerne les interconnexions verticales. Dans un premier temps, les comportements en fréquence de différents piliers de cuivre sont étudiés. Puis une étude fréquentielle ainsi que le retard induit par différent via traversant le silicium, est menée.

La dernière partie est consacrée à une des fonctionnalités possible, les inductances intégrées planaires. On étudie l'impact de différentes options technologiques, comme l'épaisseur de collage, la conductivité de silicium ou encore l'épaisseur du silicium sur le facteur de qualité. Trois règles d'optimisation sur les inductances résulteront de cette partie.

## Sommaire

<b>I</b>	<b>INTRODUCTION.....</b>	<b>90</b>
<b>II</b>	<b>LES INTERCONNEXIONS HORIZONTALES. ....</b>	<b>91</b>
II.1	ETUDE PREDICTIVE SUR LES INTERCONNEXIONS DE RDL.....	91
II.1.1	<i>Description des différents types d'interconnexion de RDL.....</i>	92
II.1.2	<i>Etude dans les domaines fréquentiel et temporel.....</i>	93
II.1.2.a	Domaine fréquentiel.....	93
II.1.2.b	Domaine temporel.....	96
II.2	ETUDES PREDICTIVES DES PERFORMANCES DES INTERCONNEXIONS DU BEOL.....	97
II.2.1	<i>Description des interconnexions du BEOL.....</i>	97
II.2.2	<i>Etudes dans les domaines fréquentiel et temporel.....</i>	98
II.2.2.a	Domaine fréquentiel.....	98
II.2.2.b	Domaine fréquentiel.....	99
II.3	REGLES DE DESSIN POUR L'OPTIMISATION DES PERFORMANCES DES INTERCONNEXIONS DU BEOL EN	
INTEGRATION 3D.....		101
II.3.1	<i>Etudes paramétriques sur l'analyse du retard de propagation et la diaphonie pour des</i>	
<i>interconnexions du BEOL lors du report de puces.....</i>		101
II.3.1.a	Empilement 3D de puces en architecture Face to Back.....	102
II.3.1.b	Empilement 3D de puces en architecture Face to Face.....	109
II.3.2	<i>Synthèse des règles de dessin.....</i>	114
<b>III</b>	<b>LES INTERCONNEXIONS VERTICALES. ....</b>	<b>116</b>
III.1	ETUDES PREDICTIVES DES PERFORMANCES DES CU-PILLARS.....	116
III.1.1	<i>Description des différents types de Cu-Pillar.....</i>	116
III.1.2	<i>Analyse fréquentielle et modèles électriques des Cu-Pillars.....</i>	117
III.2	ETUDES PREDICTIVES DES PERFORMANCES DES TSV.....	119
III.2.1	<i>Description des différents types de TSV.....</i>	120
III.2.2	<i>Analyse fréquentielle et modèles des TSV.....</i>	121
III.2.3	<i>Analyse temporelle et performances de propagation des TSV.....</i>	123
<b>IV</b>	<b>INDUCTANCES PLANAIRES .....</b>	<b>123</b>
IV.1	DESCRIPTION DE LA STRUCTURE DE REFERENCE.....	124
IV.1.1	<i>Extraction des paramètres du modèle équivalent de l'inductance.....</i>	125
IV.1.1.a	Extraction des paramètres de la partie série.....	126
IV.1.1.b	Extraction des paramètres de la partie parallèle.....	127
IV.1.1.c	Validation de la procédure d'extraction.....	128

---

IV.2	IMPACT D'UN EMPILEMENT FACE TO BACK SUR LES PERFORMANCES DE L'INDUCTANCE.....	128
IV.2.1	<i>Influence de l'épaisseur du substrat de silicium de la puce rapportée.</i> .....	128
IV.2.1.a	Description de la structure.....	129
IV.2.1.b	Résultats.....	129
IV.2.2	<i>Influence de la conductivité du silicium de la puce rapportée.</i> .....	130
IV.2.2.a	Description de la structure.....	130
IV.2.2.b	Résultats.....	131
IV.2.3	<i>Influence de l'épaisseur de collage entre les deux puces empilées.</i> .....	132
IV.2.3.a	Description de la structure.....	132
IV.2.3.b	Résultats.....	132
IV.3	ARCHITECTURE FACE TO FACE. ....	133
IV.3.1	<i>Influence de l'épaisseur et de la conductivité du substrat de silicium de la puce rapportée.</i> ....	133
IV.3.1.a	Description de la structure.....	134
IV.3.1.b	Résultat .....	134
V	CONCLUSION .....	135

## I Introduction.

Les briques élémentaires d'interconnexion constituant une chaîne complète d'interconnexion 3D peuvent se diviser en deux groupes (Figure 3-1). Le premier est nommé interconnexions horizontales, il est constitué des interconnexions de la couche de redistribution (RDL) et des interconnexions comprises dans le Back End Of Line (BEOL). Le second groupe est nommé interconnexions verticales, il est constitué des vias traversant le silicium (TSV pour Through Silicon Via) ainsi que des piliers de cuivre (Cu-Pillar).

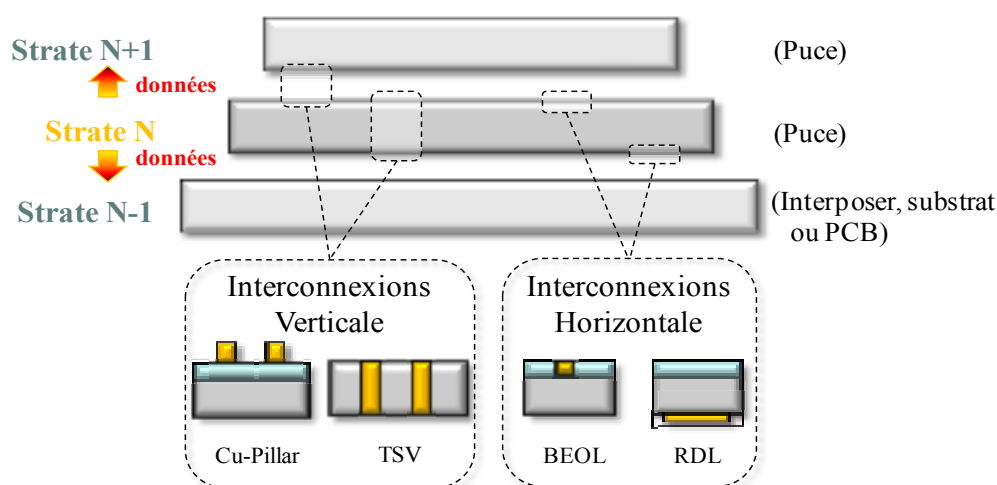


Figure 3-1 : Briques élémentaires d'interconnexions constituant une chaîne complète d'interconnexion dans un empilement 3D de puces

Dans ce chapitre sont étudiées les différentes briques élémentaires, nécessaires à la construction d'une chaîne d'interconnexions 3D performantes, afin d'arriver à une parfaite connaissance de leurs comportements dans les domaines fréquentiel et temporel. La connaissance de ces comportements est indispensable pour les chapitres suivants où ces briques élémentaires sont associées afin de prédire les performances d'une chaîne complète d'interconnexion.

Dans la première partie de ce chapitre, deux études paramétriques sont effectuées pour les interconnexions horizontales. La première concerne les interconnexions comprises dans la couche de redistribution. Les caractéristiques fréquentielles ainsi que le retard induit par ces interconnexions, en fonction de leurs longueurs et de leurs densités d'intégration, sont déterminées. La seconde étude concerne les interconnexions comprises dans le BEOL. Dans un premier temps nous exprimons les caractéristiques fréquentielles d'une interconnexion, localisée dans les niveaux semi-globaux, de génération 45 nm. Puis une étude paramétrique est proposée, donnant pour des interconnexions isolées ou couplées, les performances obtenues dans des configurations Back to Face et Face to Face en fonction de différents paramètres comme l'épaisseur de collage,



l'espacement ligne-masses, la position de dummies. Cette étude permettra d'établir six règles d'optimisation lors d'une conception en intégration 3D.

Concernant les interconnexions verticales, deux parties sont abordées. La première concerne les piliers de cuivre, Cu-Pillars. Il a été établi que les Cu-Pillar sont réalisés à l'aide de deux technologies (cf. Chapitre 1). La première est utilisée lors d'un report puce à puce et consiste à coller deux couches de cuivre avec une couche de SnAg par un phénomène de thermo-compression. La seconde est utilisée pour le report d'une puce sur un substrat, les Cu-Pillar sont composés d'une couche de cuivre et d'une couche de SnAg et un underfill comble l'espace entre les Cu-Pillar par un phénomène de capillarité. De ces deux technologies, cinq principaux éléments sont clairement identifiable par leurs densité d'intégration, un nécessaire au collage puce à puce et quatre autres nécessaires au collage puce à substrat. Une large description ainsi qu'une étude approfondie de ces éléments consistera à mieux appréhender leurs rôles et leurs positions dans une chaîne d'interconnexions, ainsi que leurs performances électriques.

La deuxième partie traite des TSV. Comme pour les Cu-Pillars, Il a été établi qu'ils sont réalisés à l'aide de technologies différentes (cf. Chapitre 1). Qu'ils soient parcellément remplis ou plein, qu'ils soient constitués avant ou après le BEOL, cinq types de TSV sont identifiables par leurs densités d'intégration et leurs dimensions. Une large description de ces éléments est exposée dans cette partie. Puis, leurs performances fréquentielles et temporelles sont déterminées.

Le chapitre se termine par une étude paramétrique réalisée sur un composant passif fonctionnel et répandu en intégration 3D l'inductance planaire. Dans un premier temps un modèle électrique et une caractérisation fréquentielle est proposée, puis l'évolution de son facteur de qualité sera étudié en fonction des options technologiques de l'intégration 3D, dans le cadre des empilements Back to Face ou Face to Face ainsi qu'en fonction de différents paramètres comme, par exemple, l'épaisseur de collage.

## **II Les interconnexions horizontales.**

Dans cette partie une large étude est menée sur les interconnexions dites horizontales, les interconnexions du BEOL et de RDL. Des études prédictives dans les domaines fréquentiel et temporel seront menées sur ces deux types d'interconnexion, afin de connaître avec exactitude leurs comportements électriques.

### **II.1 Etude prédictive sur les interconnexions de RDL.**

Les interconnexions de RDL constituent un nouveau type d'interconnexions propre à l'intégration 3D. Elles sont indispensables lors d'un assemblage multi-puces notamment pour la redistribution

des signaux d'une puce à une autre et le routage en face arrière des connectiques en provenance des TSV.

Dans cette partie, différentes interconnexions de RDL seront passées en revue en terme d'empilement technologique et de dimensions, puis une étude fréquentielle sur chaque type d'interconnexions sera effectuée en comparant leurs paramètres R, L, C et G respectifs. Pour finir, une étude temporelle sera réalisée afin de déterminer et de comparer le retard de propagation qu'elles induisent.

### II.1.1 Description des différents types d'interconnexion de RDL.

Les interconnexions de RDL étudiées sont élaborées suivant deux procédés d'intégration (Figure 3-2). Le premier (Figure 3-2-a) implique un empilement technologique composé d'un substrat de silicium (Si), d'une couche de dioxyde de silicium ( $\text{SiO}_2$  épaisseur  $3\mu\text{m}$ , permittivité  $4,2$ ) et d'une couche de résine, dans notre cas du BCB (benzocyclobutène) ayant une épaisseur de  $10\mu\text{m}$  et une permittivité relative de  $3,3$ . Le second procédé (Figure 3-2-b) implique un substrat de silicium (Si) et une couche de dioxyde de silicium ( $\text{SiO}_2$ ) d'épaisseur  $7\mu\text{m}$  et de permittivité relative de  $4,2$ . Les étapes de fabrication des deux procédés ont été détaillées dans le chapitre 1.

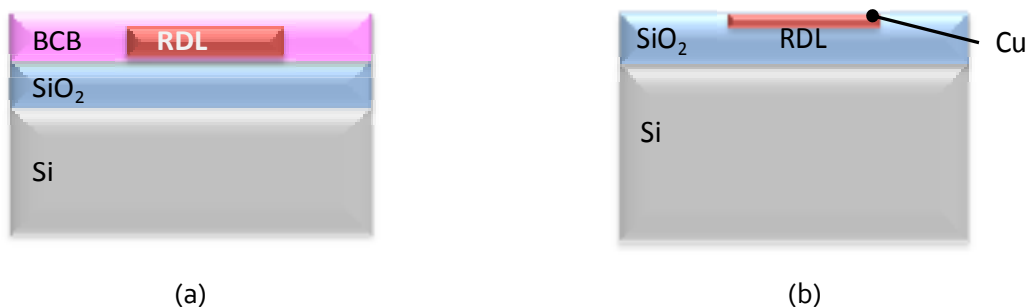


Figure 3-2 : Schéma des empilements comprenant des interconnexions de RDL.  
(a) Dans une résine épaisse. (b) Dans une cavité damascène.

Dans le but d'étudier un nombre suffisant d'interconnexions de RDL pour couvrir les nombreuses options technologiques existantes, cinq interconnexions de RDL distinctes sont choisies (

Tableau 3-1). Elles sont classées de la plus faible à la plus forte densité d'intégration. Les quatre premières sont intégrées dans de la résine épaisse et présentent des largeurs variant de  $5\mu\text{m}$  à  $30\mu\text{m}$  et des épaisseurs variant de  $1,5\mu\text{m}$  à  $7\mu\text{m}$ . Elles sont nommées en fonction de leur largeur, par exemple une interconnexion de largeur  $30\mu\text{m}$  et d'épaisseur  $7\mu\text{m}$  sera nommée RDL-w30. La dernière interconnexion étudiée est intégrée dans une cavité damascène, elle a une largeur de  $10\mu\text{m}$  (largeur importante pour pouvoir router sur de grandes longueurs), une épaisseur de  $0,5\mu\text{m}$ .

Elle est nommée en fonction de son type d'intégration, RDL-damas (pour damascène). On note  $S$ , la distance entre la ligne de RDL et le substrat.

Tableau 3-1 : Synthèse des principales informations techniques concernant les interconnexions de RDL.

	<i>RDL-w30</i>	<i>RDL-w20</i>	<i>RDL-w10</i>	<i>RDL-w5</i>	<i>RDL-damas</i>
$w_{min} (\mu m)$	30	20	10	5	10
$t (\mu m)$	7	7	2	1,5	0,5
$S (\mu m)$	3	3	3	3	0,9

## II.1.2 Etude dans les domaines fréquentiel et temporel.

### II.1.2.a Domaine fréquentiel

Plusieurs simulations électromagnétiques ont été effectuées avec le logiciel Maxwell 2D, dans le but d'obtenir les paramètres  $R$ ,  $L$ ,  $C$  et  $G$  répartis du modèle électrique d'une ligne de transmission (Cf. Chapitre2). Dans un souci de réalisme, une couche d'underfill (d'épaisseur  $40 \mu m$  et de permittivité relative 3,3) a été accolée au BCB de la structure représentée en Figure 3-2.a. Une couche de  $SiO_2$  (épaisseur  $0,9 \mu m$ ) représentant un collage direct  $SiO_2$ , suivie d'un substrat de silicium (épaisseur  $15 \mu m$ , conductivité  $15 S/m$ ) ont été accolés à l'interconnexion de RDL de la structure représentée en Figure 3-2.b. Les paramètres primaires  $R$ ,  $L$ ,  $C$  et  $G$  sont exposés dans la Figure 3-3.

L'augmentation de la résistance  $R$  des interconnexions de RDL en fonction de la fréquence est majoritairement due à l'effet de peau dans les conducteurs. Les valeurs des résistances linéiques varient en fonction de la géométrie des interconnexions. En effet, plus les dimensions sont agressives plus les résistances statiques sont élevées. L'inductance  $L$  diminue légèrement en fonction de la fréquence, ce qui indique une prédominance de l'inductance externe sur l'inductance interne.

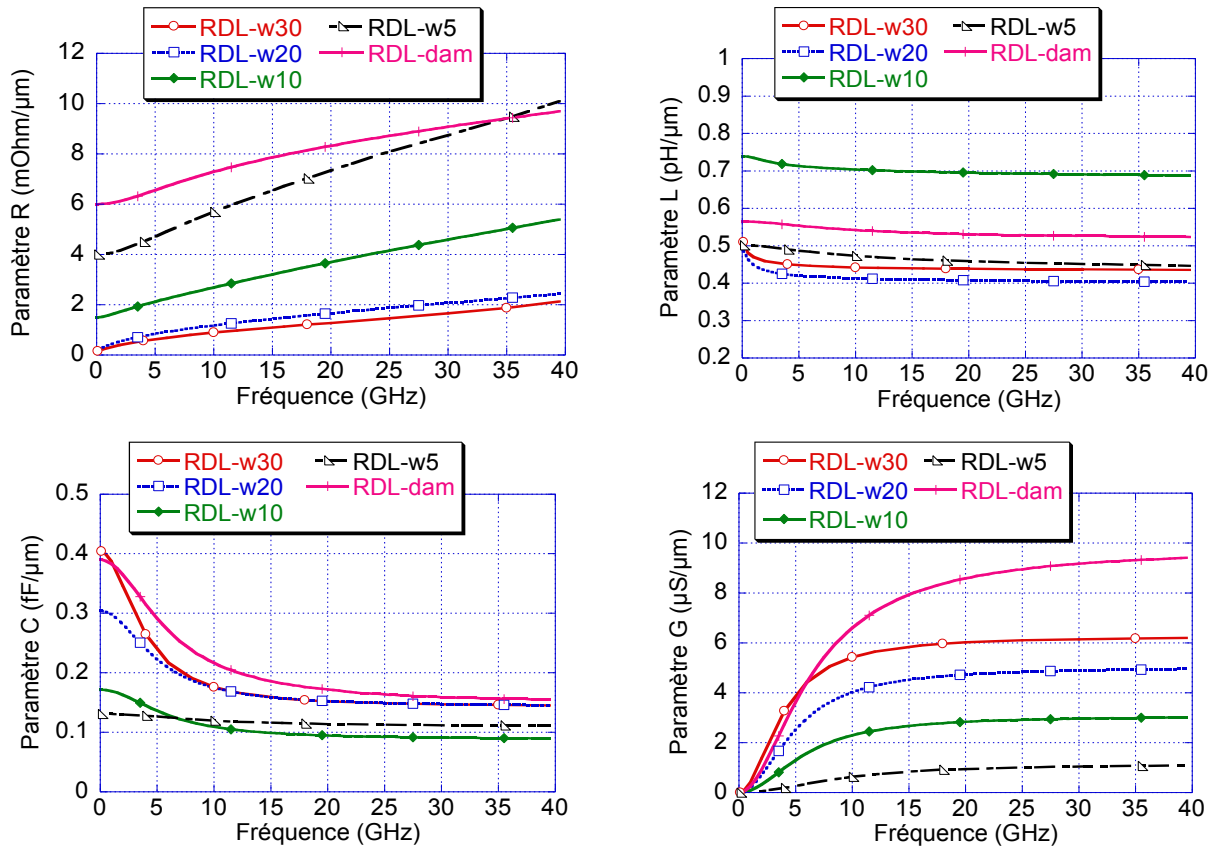


Figure 3-3 : Paramètres R, L, C et G des interconnexions de RDL.

Les capacitance et conductance C et G ne sont pas constantes sur toute la bande fréquentielle à cause de la pénétration des champs électromagnétiques dans le substrat de silicium. C'est pourquoi un modèle électrique équivalent, représentatif des phénomènes physiques, de l'admittance parallèle de la ligne de transmission est proposé en Figure 3-4.

La capacitance C présentée sur le graphique en Figure 3-3 est composée de deux termes : la capacité d'oxyde et la capacité de silicium, toutes deux représentatives des réalités physiques et elles sont constantes en fréquence.

La conductance G représente les pertes diélectriques dans le silicium.

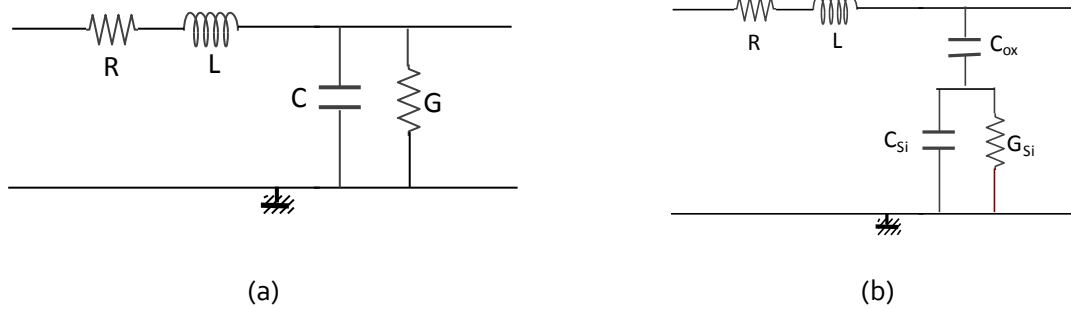


Figure 3-4 : Modèle électrique d'une interconnexion  
(a) le modèle classique (b) le modèle physique équivalent

Le

Tableau 3-2 synthétise les valeurs des paramètres R, L, C et G des différentes interconnexions de RDL. Les valeurs des résistances R, variables en fréquence à cause de l'effet de peau, sont reportées en statique et à 20 GHz. Pour les inductances L, quasiment constantes, les valeurs statiques sont également rapportées.  $C_{ox}$ ,  $C_{si}$  et  $G_{si}$  sont déterminés à l'aide du modèle équivalent donné en figure 3-4. En effet en exprimant le modèle classique et le modèle équivalent sous la forme d'admittance, respectivement  $Y_{classique}$  (Équation 3-1) et  $Y_{équivalent}$  (Équation 3-2) et en procédant par identification (Équation 3-3 et Équation 3-4) on obtient les paramètres  $C_{ox}$ ,  $C_{si}$  et  $G_{si}$  en fonction de C et G (Équation 3-5 à Équation 3-7).

$$Y_{classique} = G + jC\omega \quad \text{Équation 3-1}$$

$$Y_{équivalent} = \frac{\omega^2 \cdot G_{si}^2 \cdot C_{ox}^2}{G_{si}^2 + \omega^2(C_{ox} + C_{si})^2} + j\omega \frac{C_{ox}G_{si} + \omega^2 C_{ox}C_{si}(C_{ox} + C_{si})}{G_{si}^2 + \omega^2(C_{ox} + C_{si})^2} \quad \text{Équation 3-2}$$

$$G = \frac{\omega^2 \cdot G_{si}^2 \cdot C_{ox}^2}{G_{si}^2 + \omega^2(C_{ox} + C_{si})^2} \quad \text{Équation 3-3}$$

$$C = \frac{C_{ox}G_{si} + \omega^2 C_{ox}C_{si}(C_{ox} + C_{si})}{G_{si}^2 + \omega^2(C_{ox} + C_{si})^2} \quad \text{Équation 3-4}$$

$$\text{A basse fréquence } (\omega \rightarrow 0); \quad C_{ox} = C \quad \text{Équation 3-5}$$

$$\text{A haute fréquence } (\omega \rightarrow \infty); \quad C_{si} = \frac{C_{ox}C}{C_{ox} - C} \quad \text{Équation 3-6}$$

$$\text{A haute fréquence } (\omega \rightarrow \infty); \quad G_{si} = \frac{G(C_{ox} + C_{si})^2}{C_{ox}^2} \quad \text{Équation 3-7}$$

Tableau 3-2 : Valeurs des paramètres R, L, C et G des interconnexions de RDL.

	$R \text{ (m}\Omega/\mu\text{m)}$	$R \text{ à } 20\text{GHz}$	$L \text{ (pH}/\mu\text{m)}$	$C_{ox} \text{ (fF}/\mu\text{m)}$	$C_{si} \text{ (fF}/\mu\text{m)}$	$G_{si} \text{ (}\mu\text{S}/\mu\text{m)}$
			$(\text{m}\Omega/\mu\text{m})$			
<b>RDL-w30</b>	0,145	1,2	0,5	0,404	0,22	14,8
<b>RDL-w20</b>	0,214	1,6	0,48	0,304	0,28	18,2
<b>RDL-w10</b>	1.5	3,3	0,72	0,17	0,18	12,7
<b>RDL-w5</b>	4	7,3	0,5	0,13	0,8	55,7
<b>RDL-damas</b>	6	8,3	0,56	0,39	0,25	25,3

Une vérification analytique sur la résistance R (Équation 3-8), nous confirme que les simulations effectuées sont en concordance avec la théorie. Les valeurs théoriques obtenues sont synthétisées dans le

Tableau 3-3.

$$R_1 = R_{\text{ligne}} * \frac{3}{2} = \rho \frac{1}{\omega * t} * \frac{3}{2} \quad \text{Équation 3-8}$$

Remarque : Etant donné que les rubans de masse sont de la même dimension que le ruban de ligne, le coefficient 3/2 correspond au retour de courant dans les masses.

Tableau 3-3 : Valeurs théoriques des résistances R en statique.

	<i>RDL-w30</i>	<i>RDL-w20</i>	<i>RDL-w10</i>	<i>RDL-w5</i>	<i>RDL-damas</i>
$R_{\text{théorique}}$ (mΩ/μm)	0,142	0,214	1,5	4	6

### II.1.2.b Domaine temporel

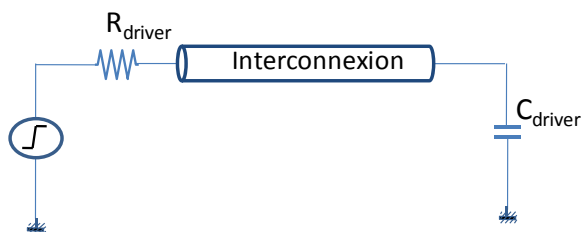
Afin de visualiser le retard généré par chaque interconnexion de RDL, indépendamment des conditions de charge, on charge par les mêmes drivers toutes ces interconnexions. Les drivers utilisés sont typiques de ceux de la technologie 65 nm. Ils sont modélisés par une résistance ( $R_{\text{driver}}$ ) en entrée de ligne et par une capacité ( $C_{\text{driver}}$ ) en sortie (Figure 3-5-a). Lesquels sont déterminés par les équations suivantes :

$$R_{\text{driver}} = R_{\text{acc}} / W_g = 170 / 6.8 = 25 \Omega \quad \text{Équation 3-9}$$

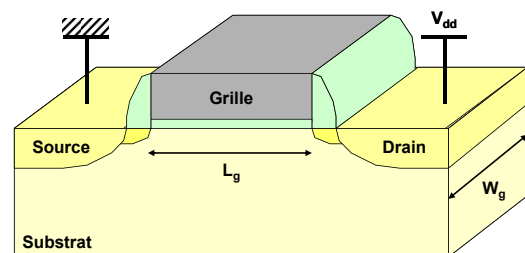
$$C_{\text{driver}} = C_{\text{inv}} * W_g = 3.45 * 6.8 = 23.2 \text{ fF} \quad \text{Équation 3-10}$$

$R_{\text{driver}}$  correspond à l'ensemble des contributions de la résistance d'accès de l'inverseur (constitué de deux transistors) situé en entrée ( $R_{\text{acc}}$ ), multiplié à la largeur de grille des transistors ( $W_g$ ) (Figure 3-5-b), alors que  $C_{\text{driver}}$  divise la capacité de charge de l'inverseur située en sortie ( $C_{\text{inv}}$ ) par la largeur de grille ( $W_g$ ).

Pour les études présentées,  $R_{\text{acc}} = 170 \Omega$ ,  $C_{\text{inv}} = 3.45 \text{ fF}/\mu\text{m}$  et  $W_g$  correspond à la largeur de grille des transistor type NMOS, soit  $6.8 \mu\text{m}$ .



(a)



(b)

Figure 3-5 : (a) Modélisation des drivers d'entrée et de sortie (b) et topologie des transistors

La Figure 3-6, représente le retard induit par chaque type d'interconnexion de RDL en fonction de leur longueur. On remarque que les interconnexions de RDL intégrées dans de la résine génèrent un retard plus faible, principalement pour les RDL-w5 et RDL-w10. Bien que leurs résistances soient élevées, elles ont une capacité d'oxyde nettement plus faible que les autres. Ce qui confirme qu'en augmentant les densités d'intégration on diminue le retard des interconnexions.

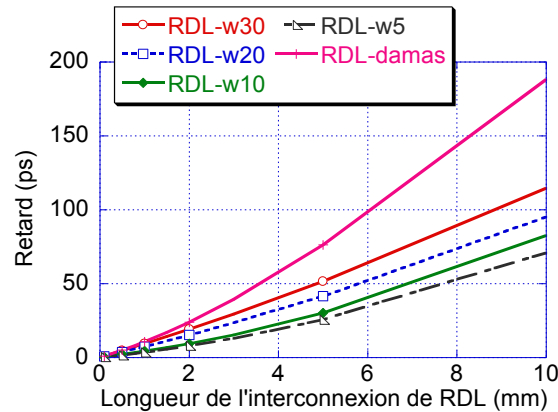


Figure 3-6 : Retard induit par les différentes interconnexions de RDL en fonction de leur longueur.

## II.2 Etudes prédictives des performances des interconnexions du BEOL.

### II.2.1 Description des interconnexions du BEOL.

Lors d'une intégration 3D les interconnexions du BEOL utilisées pour le routage des signaux sont situées dans les derniers niveaux de métallisation, les plus épais.

Une interconnexion située dans les niveaux de métallisation globaux intégrée dans le BEOL d'une puce est schématisée en figure 3-7. Son empilement technologique est composé d'un substrat de silicium (Si) et d'une couche de dioxyde de silicium ( $\text{SiO}_2$ ) d'épaisseur  $4,65 \mu\text{m}$ , comprenant les conducteurs d'épaisseurs  $850 \text{ nm}$ .

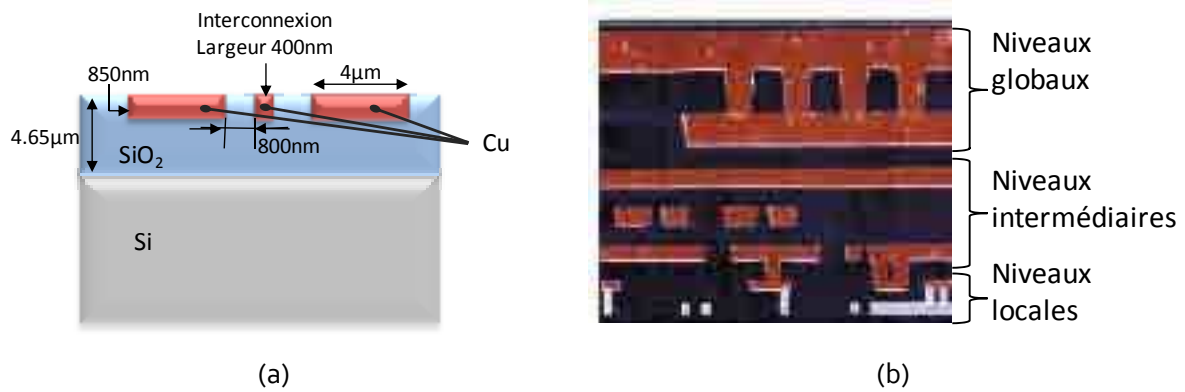


Figure 3-7 : (a) Vue schématique d'une interconnexion de niveau global.  
(b) Vue en coupe d'un empilement d'interconnexions.

La ligne signal et les rubans de masse sont respectivement de largeur 400 nm et 4 μm. La distance qui les sépare est de 800 nm (Figure 3-7-a).

## II.2.2 Etudes dans les domaines fréquentiel et temporel.

### II.2.2.a Domaine fréquentiel

Afin d'obtenir par simulations électromagnétiques les paramètres R, L, C et G linéiques du modèle électrique de la ligne de transmission, représentés en Figure 3-8, une couche d'underfill (d'épaisseur 40 μm et de permittivité 3,3) a été accolée à la couche de SiO<sub>2</sub> comportant l'interconnexion d'BEOL.

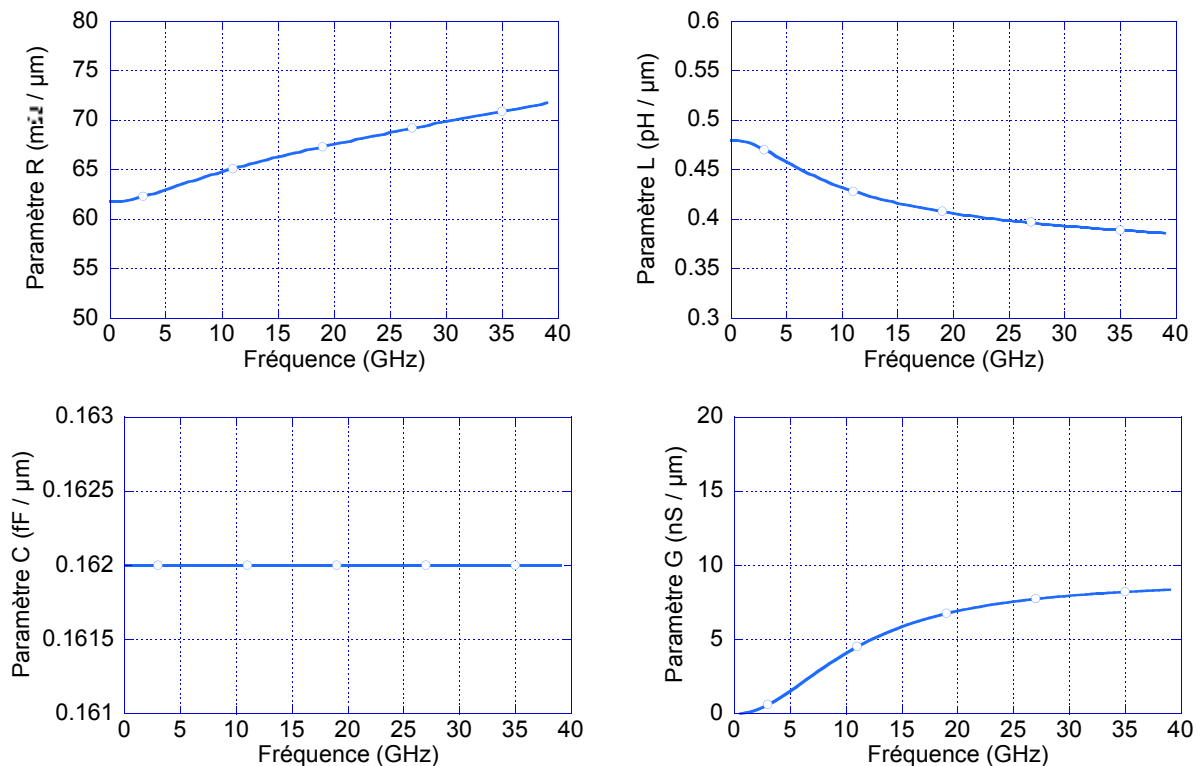




Figure 3-8 : Paramètres R, L, C et G d'une interconnexion du BEOL.

L'augmentation de la résistance R est due à l'effet de peau dans les conducteurs. Les dimensions de cette interconnexion, bien que relâchées car située dans les niveaux globaux, restent bien plus faibles que celles des interconnexions de RDL. Elles impliquent une résistance statique élevée, de  $61 \text{ m}\Omega/\mu\text{m}$ , par rapport à une interconnexion de RDL fortement intégrée (RDL-w5;  $R_{\text{statique}} = 5 \text{ m}\Omega/\mu\text{m}$ ).

L'inductance L diminue légèrement en fonction de la fréquence. La capacitance C est constante et la conductance G n'excède pas  $10 \text{ nS}/\mu\text{m}$  ce qui la rend négligeable. Ceci nous indique que les champs électromagnétiques ne pénètrent pas dans le substrat de silicium. En effet, dans cette configuration précise, les champs sont confinés dans l'oxyde  $\text{SiO}_2$ .

Ceci nous permet d'extraire les éléments caractéristiques de cette interconnexion en utilisant le modèle classique des lignes de transmissions. Les paramètres R, L, C et G sont extraits à partir des graphiques donnés en Figure 3-8, où R et L sont définies à partir des valeurs statiques. La capacitance C est constante et la conductance G est considérée comme nulle, car sa valeur est si faible qu'elle n'est pas représentative d'un phénomène physique. La synthèse des paramètres R, L, C de l'interconnexion est donnée dans le Tableau 3-4.

Tableau 3-4 : Valeurs des paramètres R, L, C de l'interconnexion de BEOL

$R \text{ (m}\Omega/\mu\text{m)}$	$L \text{ (pH}/\mu\text{m)}$	$C \text{ (fF}/\mu\text{m)}$
61	0,48	0,162

### II.2.2.b *Domaine temporel*

Dans le but de comparer les deux familles d'interconnexions horizontales (RDL et BEOL), on utilise les drivers présentés au paragraphe II.1.2.b. La Figure 3-9 représente le retard induit par l'interconnexion de BEOL, en fonction de sa longueur.

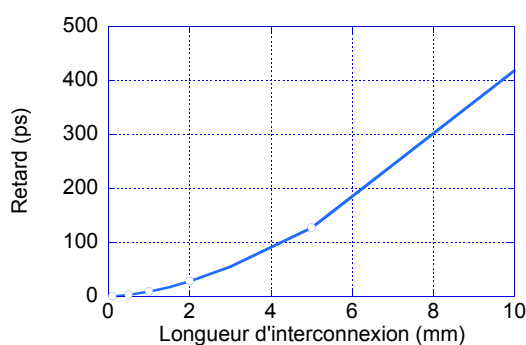


Figure 3-9 : Retard provoqué par une interconnexion de BEOL en fonction de sa longueur.

La longueur 10mm n'est pas une longueur représentative des interconnexions présentes dans le BEOL de circuit intégré mais cela permet de confirmer la forte augmentation du retard en fonction de la longueur d'interconnexion.

La comparaison de ce retard avec celui engendré par les interconnexions de RDL est montrée Figure 3-10. La Figure 3-10-a représente les retards des interconnexions de RDL et de BEOL pour des longueurs pouvant aller jusqu'à 10 mm. On remarque que l'interconnexion du BEOL implique un retard beaucoup plus important que celui de RDL et ce d'autant plus pour des longueurs très élevées. La Figure 3-10-b est un agrandissement du graphique précédent où l'on ne s'intéresse qu'à des longueurs inférieures à 1,5 mm.

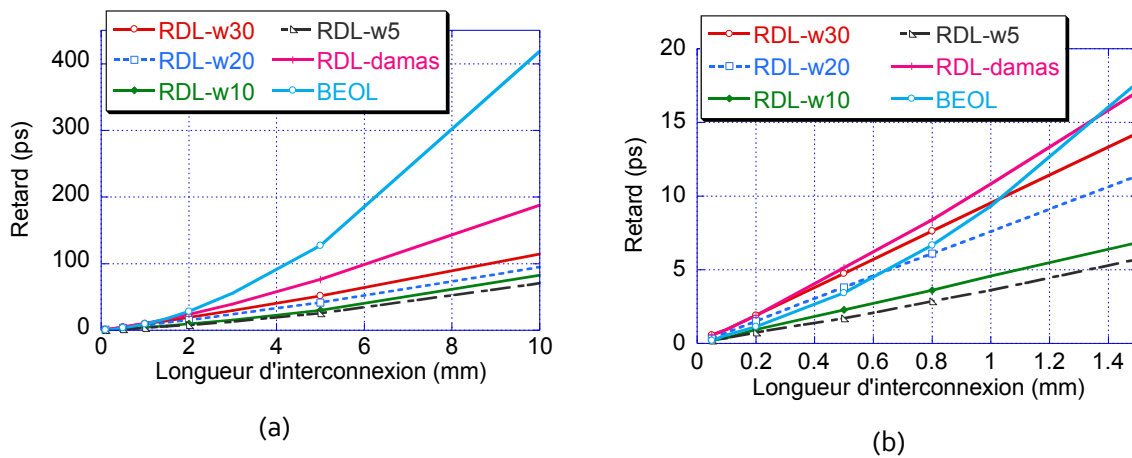


Figure 3-10 : Confrontation des interconnexions de RDL et de BEOL (a) Pour des longueurs d'interconnexion allant jusqu'à 10mm (b) Pour des longueurs d'interconnexion allant jusqu'à 1,5mm

Pour des longueurs d'interconnexion supérieures à 1,4 mm les interconnexions de RDL sont plus performantes que les interconnexions du BEOL. Cependant, ici tout est une question de compromis, l'interconnexion du BEOL peut présenter de meilleures performances qu'une interconnexion de RDL, principalement pour des faibles longueurs d'interconnexion. En effet, pour une longueur inférieure ou égale à 800nm l'interconnexion du BEOL engendre moins de retard que trois types d'interconnexion de RDL (RDL-damas, RDL-w30 et RDL-w20).

Les interconnexions du BEOL des niveaux locaux et intermédiaires ont été très étudiées dans la littérature mais l'originalité de cette étude consiste en l'étude des lignes semi-globales considérées pour le routage en trois dimensions. Il reste à déterminer l'influence d'un empilement de type 3D sur leurs performances. La partie suivante est une étude paramétrique sur les performances des interconnexions de BEOL dans un environnement 3D, dans le but de déterminer des règles d'optimisation.

### II.3 Règles de dessin pour l'optimisation des performances des interconnexions du BEOL en intégration 3D.

Dans cette partie, les performances dans le domaine temporel des interconnexions situées dans le BEOL sont étudiées, dans le but de quantifier l'impact du type d'empilement de deux puces sur le retard. Par le biais d'une étude paramétrique, les performances pour des interconnexions isolées ou couplées, sont obtenues dans des configurations Back to Face et Face to Face en fonction de différents paramètres comme l'épaisseur de collage, l'espacement ligne-masses, la position des dummies. A partir de ces études, plusieurs critères pourront être définis dans le but de diminuer le retard causé par les interconnexions du BEOL. De ce fait le retard à 50 % sera analysé pour les interconnexions isolées ainsi que le niveau de diaphonie pour les interconnexions couplées. Pour ces études nous supposons la transmission de signaux quasi-indiciels avec un temps de montée de 40 ps.

#### II.3.1 Etudes paramétriques sur l'analyse du retard de propagation et la diaphonie pour des interconnexions du BEOL lors du report de puces

L'empilement technologique de la puce inférieure reste le même tout au long de cette étude (Figure 3-11-a). Il est composé d'un substrat de silicium de conductivité 10 S/m, d'une couche mince d'isolant à très faible permittivité (Ultra Low K ULK ;  $\epsilon_r = 2,6$ ) et d'une couche de dioxyde de silicium ( $\text{SiO}_2$  ;  $\epsilon_r = 4,2$ ) englobant les conducteurs. Les dimensions de ces couches métalliques correspondent aux valeurs limites de la technologie 45 nm, l'épaisseur est de 900 nm et les largeurs sont de 400 nm pour la ligne signal et 3  $\mu\text{m}$  pour les masses (Figure 3-11-b).

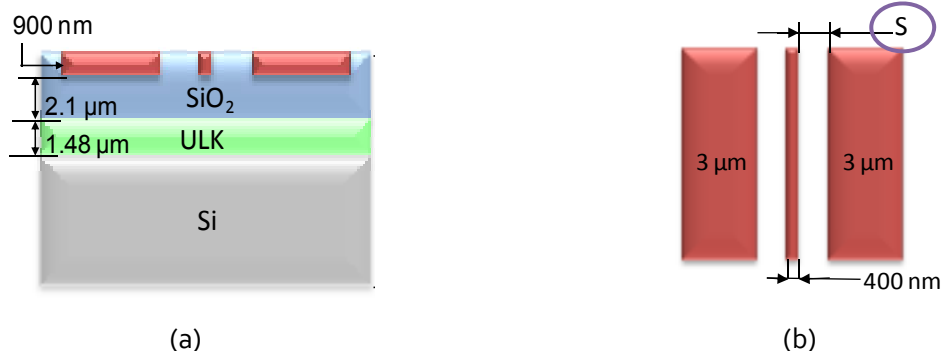


Figure 3-11 : Vue schématique de la puce inférieure. (a) Empilement technologique. (b) Topologie des conducteurs.

Le distance entre la ligne signal et les métallisations de masse est notée S. Suivant les études, cette distance S varie de 0,8 à 1,5  $\mu\text{m}$ .

### II.3.1.a Empilement 3D de puces en architecture Face to Back.

Le substrat de silicium de la puce supérieure est collé au BEOL de la puce inférieure, au moyen d'un collage direct  $\text{SiO}_2$  (Figure 3-12). L'épaisseur du collage, notée  $d$ , peut varier de 0,1 à 6  $\mu\text{m}$  selon la technologie employée.

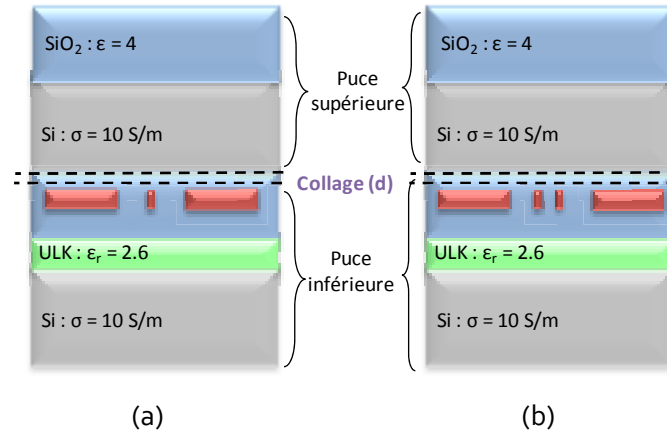


Figure 3-12 : Vue schématique d'un empilement Face to Back. (a) interconnexion isolée  
(b) interconnexions couplées.

#### II.3.1.a.1 Interconnexion isolée du BEOL en empilement Face to Back

Dans le but de déterminer le retard induit par l'interconnexion dans son nouvel environnement, l'influence sur le retard de propagation à 50% de plusieurs paramètres comme l'épaisseur de collage ( $\text{SiO}_2$ ;  $\epsilon_r = 4,2$ ) entre les deux puces, l'espacement entre les conducteurs et le design de l'interconnexion (CPW coplanar wave guide, MSW microstrip wave guide) a été étudiée.

##### ➤ Influence du report de la puce supérieure

Afin d'évaluer l'influence du report de la puce supérieure sur le retard lié aux interconnexions de la puce inférieure, des simulations ont été réalisées suivant deux empilements technologiques :

- Puce inférieure + épaisseur de collage, notée « Sans puce supérieure ».
- Puce inférieure + épaisseur de collage + puce supérieure, notée « Avec puce supérieure ».

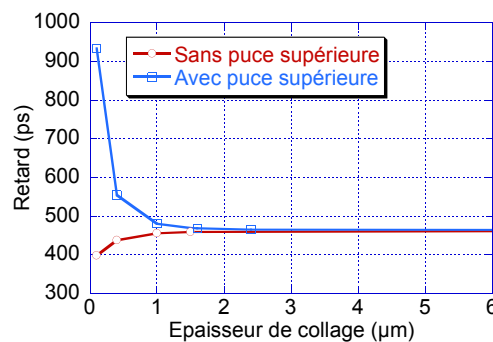


Figure 3-13 : Retard de propagation sur une interconnexion du BEOL en fonction de l'épaisseur de collage avec et sans puce supérieure.

Comme le montre la figure 3-11 pour de faibles épaisseurs de collage, l'ajout d'une puce supérieure dégrade fortement le retard. Le silicium est très proche des conducteurs constituant l'interconnexion, les champs se propagent dans le SiO<sub>2</sub> et notablement dans le silicium, ce qui engendre des pertes et donc des temps retard qui augmentent lorsque l'épaisseur de collage diminue. Lorsque qu'il n'y a pas de puce supérieure le phénomène est inversé, les champs se propagent majoritairement dans le SiO<sub>2</sub> et dans l'air. Les temps de retard augmentent avec l'épaisseur de SiO<sub>2</sub>. On observe entre les deux types d'empilements un écart relatif sur le retard de 135% pour une épaisseur de collage de 100 nm. Cependant, cet écart s'estompe rapidement puisque les retards se stabilisent lorsque l'épaisseur de collage atteint 1,6 µm. Cela signifie qu'à partir de cette valeur, l'épaisseur de collage est suffisante pour masquer les effets provoqués par tout ce qui se trouve au-delà et le report de la puce supérieure n'a plus d'influence.

Etant donné que le retard se stabilise pour une épaisseur de collage de 1,6 µm, et que l'espacement entre les conducteurs est de 800 nm, on peut émettre une loi empirique liant l'épaisseur de collage et l'espacement entre la ligne signal et les conducteurs de masse.

Cette loi empirique est la suivante (Équation 3-11) : Il existe une épaisseur de collage seuil ( $d_{seuil}$ ), qui est égale à deux fois l'espacement entre les conducteurs ( $S$ , distance entre la ligne de signal et une ligne de masse), à partir de laquelle la puce supérieure n'a plus d'impact sur le retard.

$$d_{seuil} = 2 \cdot S$$

Équation 3-11

Critère d'optimisation n°1 : L'augmentation de l'épaisseur de collage permet de diminuer le retard à 50%. Afin de s'affranchir de l'influence du substrat rapporté d'un empilement Face to Back, l'épaisseur de collage doit être au moins de l'ordre du double de la distance entre les métallisations d'un même niveau métal pour le design d'une interconnexion de type CPW.

Par exemple, pour des métallisations espacées de 800 nm, le fait de passer d'une épaisseur de collage de 100 nm à 200 nm (zone où le substrat exerce une influence), permet de réduire le retard de 17 %.

- Influence de l'espacement entre les conducteurs d'une interconnexion CPW du BEOL pour un empilement Face to Back

Dans un environnement 3D Face to Back, l'influence de l'espacement entre les conducteurs ( $S$ ) constituant une interconnexion de topologie CPW sur le retard est évaluée.

Trois interconnexions du BEOL ayant un espacement  $S$  différent sont étudiées :

- $S = 1500$  nm.
- $S = 1200$  nm.

- $S = 800 \text{ nm}$ .

La comparaison des temps de retard à 50% sur ces trois interconnexions (Figure 3-14) permet, dans un premier temps, de valider la relation entre l'espacement des conducteurs et la valeur de l'épaisseur de collage pour laquelle le substrat de la puce rapportée n'a plus d'influence (Critère d'optimisation n°1) et dans un second temps, d'évaluer l'influence de cet espacement sur le retard à 50%. La figure 3-12 présente les résultats obtenus.

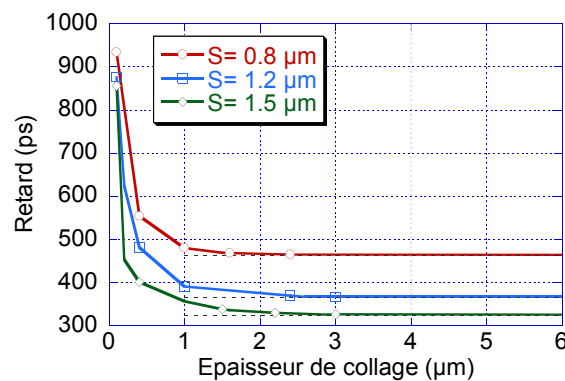


Figure 3-14 : Retard en fonction de l'épaisseur de collage entre les deux puces empilées en Face to Back pour différents espacements entre les métallisations d'une interconnexion CPW du BEOL.

Les résultats présentés ci-dessus valident l'hypothèse de l'Équation 3-11:

- Pour  $S = 0.8 \text{ μm}$   $\rightarrow d_{\text{seuil}} \approx 1.6 \text{ μm}$
- Pour  $S = 1.2 \text{ μm}$   $\rightarrow d_{\text{seuil}} \approx 2.4 \text{ μm}$
- Pour  $S = 1.5 \text{ μm}$   $\rightarrow d_{\text{seuil}} \approx 3 \text{ μm}$

Nous remarquons que le retard devient moins important lorsque  $S$  augmente. Ceci vient du fait que l'augmentation de la distance entre les conducteurs entraîne une diminution de la valeur de la capacité  $C$  et une augmentation de la valeur de l'inductance  $L$ .

Critère d'optimisation n°2 : Dans une configuration Face to Back, afin de minimiser le retard de propagation à 50%, l'espacement entre les métallisations de l'interconnexion de BEOL de type CPW doit être deux fois plus petit que l'épaisseur de collage, allié à un espacement entre les métallisations de l'interconnexion de BEOL de type CPW le plus important possible.

Par exemple, pour une épaisseur de collage équivalente ( $d = 200 \text{ nm}$ ) un espacement entre les métallisations de  $1,2 \text{ μm}$  à la place d'un espacement de  $0,8 \text{ μm}$ , permet de diminuer le retard de 11 %.

- Influence de la topologie de l'interconnexion du BEOL

L'influence sur le retard, de deux interconnexions de topologies différentes, en fonction de l'épaisseur de collage est étudiée :

- Topologie CPW, coplanaire.
- Topologie MSW, microruban.

Dans le cas de la structure coplanaire, les lignes conductrices sont situées au niveau de métallisation M7 (Figure 3-12), alors que pour la structure microruban, la ligne centrale est au niveau M7 et un plan de masse ce trouve au niveau M5 (Figure 3-15-a).

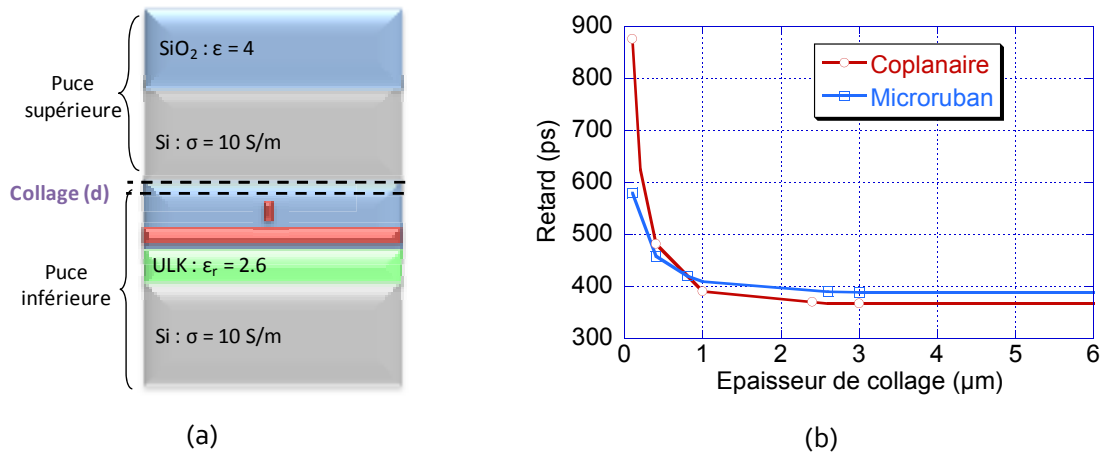


Figure 3-15 : Influence des topologies des interconnexions du BEOL en empilement Back to Face sur le retard. (a) Vue schématique de la topologie CPW. (b) Retard en fonction de l'épaisseur de collage pour des structures CPW et MSW.

Pour de faibles épaisseurs de collage, inférieures à 0,6  $\mu\text{m}$ , le retard est plus faible avec la structure microruban. Pour des épaisseurs de collage supérieures, le retard est plus faible avec la structure coplanaire. A partir d'une épaisseur de collage de 3  $\mu\text{m}$  le retard devient indépendant de l'épaisseur de collage et on obtient un écart relatif de 5,5 % entre les deux topologies d'interconnexion.

Critère d'optimisation n°3 : Il est possible d'adapter la topologie des interconnexions, soit avec une topologie CPW, soit avec une topologie MSW (si les contraintes de fabrication le permettent), pour minimiser le retard à 50 % dans un empilement Face to Back.

Par exemple, pour une épaisseur de collage de 100 nm, une topologie MSW permet de diminuer le retard de 34 %, alors que pour une épaisseur de collage de 1  $\mu\text{m}$  c'est la topologie CPW qui permet de diminuer le retard.

### II.3.1.a.2 Interconnexions couplées dans le BEOL

Dans cette partie, nous allons étudier l'influence du report d'une puce sur la diaphonie et le retard de deux interconnexions couplées dans le BEOL de la puce inférieure, en configuration Face to Back.

Pour cette étude nous supposons que les signaux activés sur les interconnexions sont de types quasi-indiciels, d'amplitude 1 V et de temps de montée 40 ps. Ils mettent en jeu des fronts montants et descendants qui nous permettent d'étudier les temps de retard à 50% des interconnexions couplées. Nous considérons trois types de signaux délivrés par les drivers d'entrée des interconnexions. Le premier étant un front montant, le second un front descendant et le dernier sans tension (Figure 3-16).

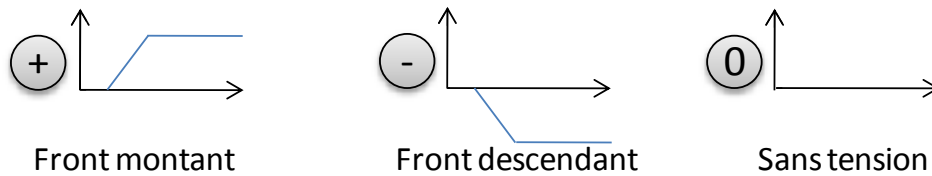


Figure 3-16 : Signaux délivrés par les drivers d'entrée des interconnexions

La notation « excitation en phase » correspond au cas où les interconnexions sont excitées par le même type de signal. La notation « excitation en antiphasé » correspond au cas où les interconnexions sont excitées par des signaux ayant des fronts opposés. La notation « excitation seule » correspond au cas où l'une des interconnexions n'est pas excitée (Figure 3-17).

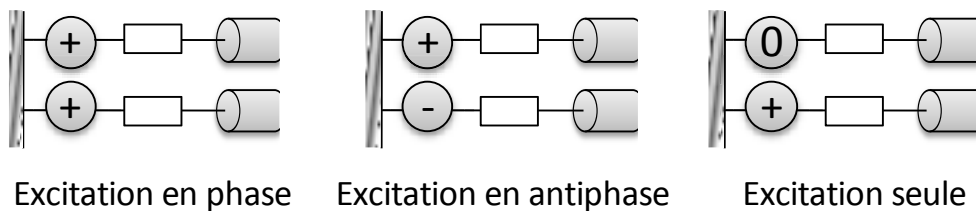


Figure 3-17 : Définition des conditions d'excitation (activation).

➤ Influence de l'espacement entre les conducteurs pour une interconnexion CPW du BEOL

L'étude porte sur une interconnexion coplanaire CPW intégrée au niveau M7 de la puce inférieure. Deux interconnexions ayant un espacement ligne-masse (S) différent sont étudiées:

- Interconnexion CPW avec  $S = 0,8 \mu\text{m}$ .
- Interconnexion avec  $S = 1,2 \mu\text{m}$ .

La figure 3-16 présente les retards à 50% sur les interconnexions CPW couplées en fonction des trois types d'activation des signaux sur celles-ci, ainsi que le niveau de diaphonie (crosstalk) obtenu dans le cas d'une « excitation seule » sur l'interconnexion non activée (niveau de signal parasite normalisé sur l'interconnexion au repos). Ces résultats sont donnés en fonction de l'épaisseur de collage et pour les deux valeurs d'espacement S entre les métallisations des interconnexions CPW.



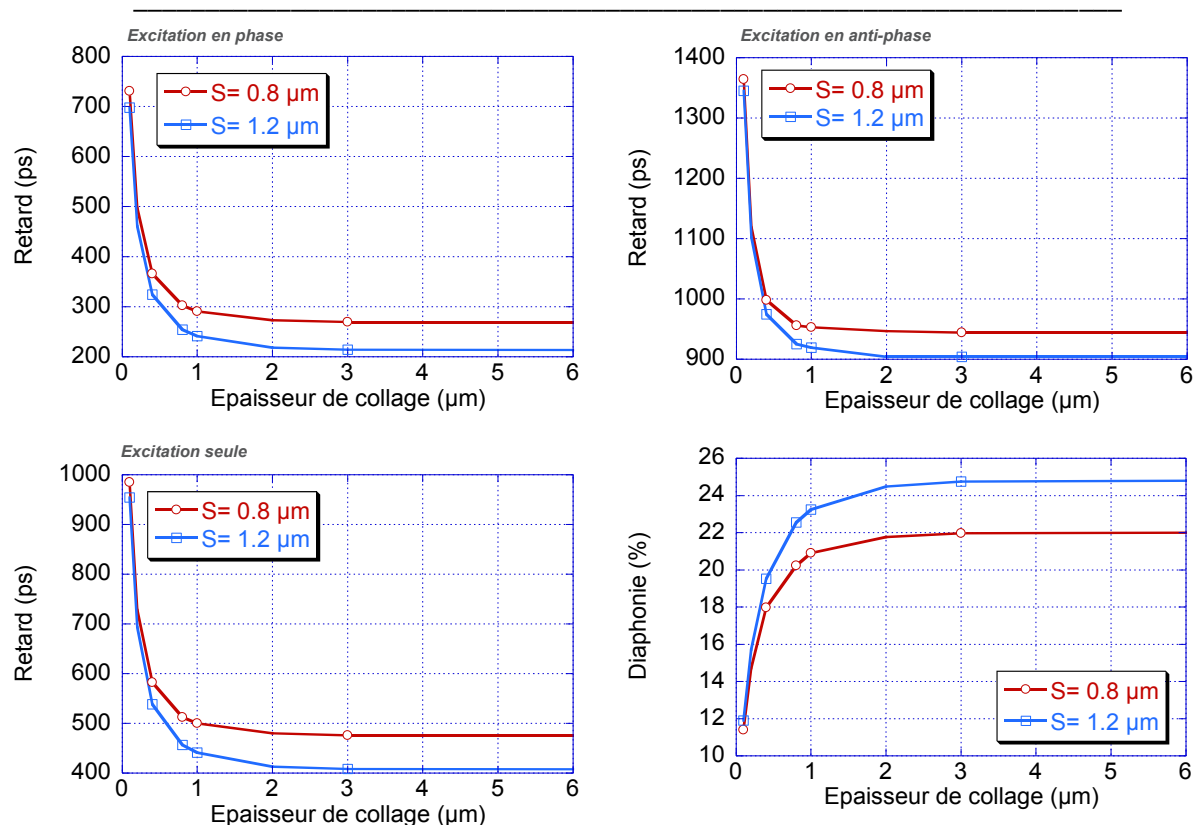


Figure 3-18 : Temps de retard pour chaque type d'excitation et diaphonie en fonction de l'épaisseur de collage et pour deux espacements  $S$  différents dans la topologie CPW.

De la même manière que pour une interconnexion isolée on remarque que plus l'espacement ligne-masse est grand, plus le retard diminue, cependant la diaphonie augmente comme cela était prévisible.

Quelque soit le type d'excitation, le retard augmente considérablement pour de faibles épaisseurs de collage ( $< 3 \mu\text{m}$ ), alors que la diaphonie diminue.

Un compromis est donc à trouver entre les performances établies en termes de retard à 50 % et la diaphonie. En effet, une optimisation du retard qui passerait par une augmentation de l'épaisseur de collage ou une augmentation de l'espacement entre les conducteurs, impliquerait une augmentation de la diaphonie et inversement.

### ➤ Influence de la topologie des interconnexions couplées du BEOL pour un empilement Face to Back

Dans cette partie, nous allons étudier l'influence des deux topologies sur le temps de propagation (Excitation en phase, en anti-phase et seule) en fonction de l'épaisseur de collage.

- Topologie CPW, coplanaire.
- Topologie MSW, microruban.

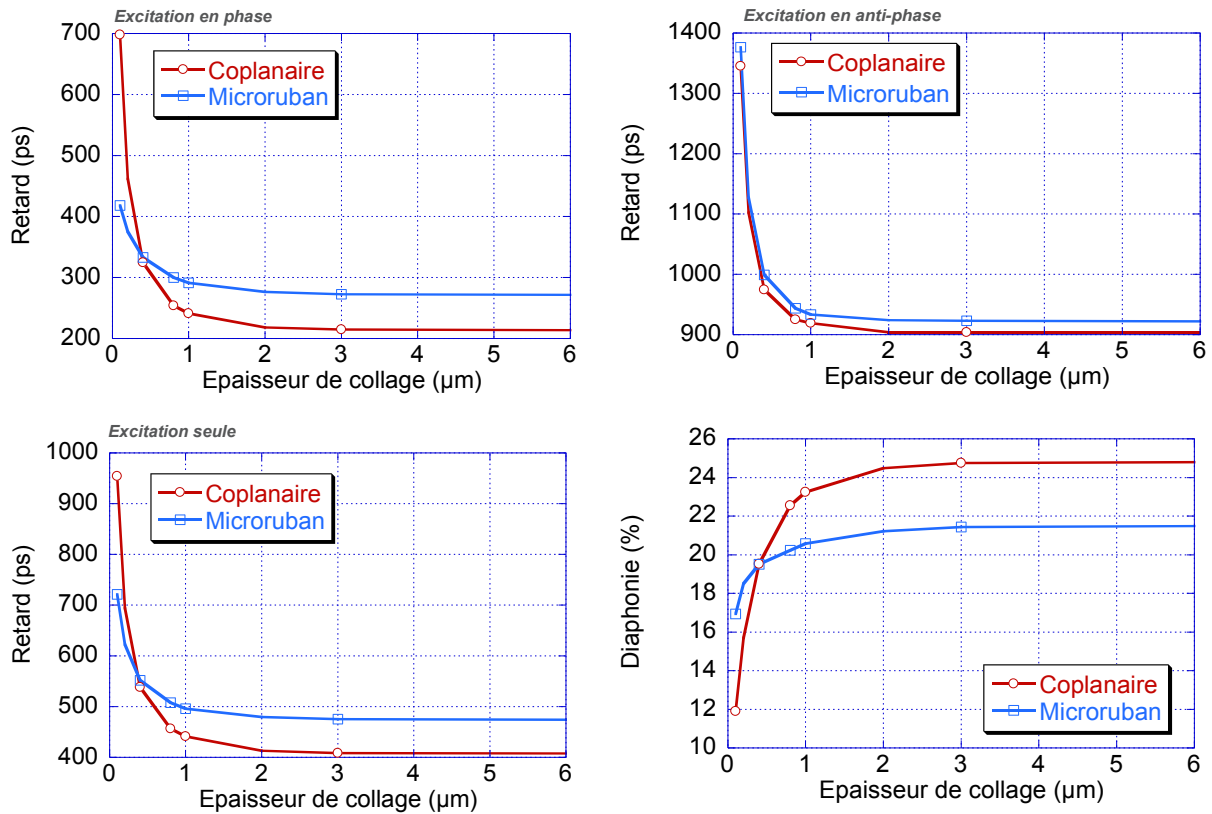


Figure 3-19 : Temps de retard pour chaque type d'excitation et diaphonie en fonction de l'épaisseur de collage et pour les deux topologies CPW et MSW des interconnexions

La topologie MSW microruban, par rapport à la topologie CPW coplanaire, génère un retard plus faible et une diaphonie plus élevée pour de faibles épaisseurs de collage. Lorsque l'épaisseur de collage dépasse 800 nm, l'effet est inversé. Ce commentaire n'est valable que lors d'une excitation en phase et d'une excitation seule car, dans le cas d'une excitation en anti-phase, malgré des valeurs de retard très fortes, l'architecture coplanaire est celle qui donne de meilleurs résultats quelque soit l'épaisseur de collage.

A partir de ces résultats un compromis entre les performances définies en termes de retard et de diaphonie doit être appliqué. En effet une optimisation du retard passerait par une augmentation de l'épaisseur de collage ou un changement de design, ce qui impliquerait une augmentation de la diaphonie et inversement.

Critère d'optimisation n°4 : Pour des interconnexions du BEOL couplées il existe toujours un compromis pour minimiser soit le retard à 50% soit la diaphonie, en jouant soit sur le type de topologie d'interconnexion (CPW ou MSW), soit sur l'espacement entre métallisation pour la topologie CPW.

Par exemple, passer d'un espacement entre les métallisations de 800 nm à 1200 nm, pour une épaisseur de collage de 1  $\mu\text{m}$ , permet de diminuer le retard de 12 % (lors d'une excitation seule) mais

augmente la diaphonie de 2,2 %. De plus pour une épaisseur de collage de 100 nm, une topologie MSW, diminue le retard de 23 % mais augmente la diaphonie de 4,5 %.

### **II.3.1.b** *Empilement 3D de puces en architecture Face to Face*

Nous nous intéressons maintenant au cas où le BEOL de la puce supérieure est directement collé au BEOL de la puce inférieure (empilement Face to Face), au moyen d'un collage direct  $\text{SiO}_2$ . A ce niveau, il est indispensable de prendre en compte les effets d'un des éléments technologiques incontournable dans la fabrication des réseaux d'interconnexions dans le BEOL, les « dummies ». En effet, le procédé d'intégration « double damascène », utilisé depuis l'intégration du cuivre à la place de l'aluminium, nécessite une étape de polissage mécanique et chimique appelée CMP (Chemical Mechanical Polishishing). Pour que cette étape de polissage soit réalisée convenablement un certain nombre de règles doit être respecté, comme le respect d'une largeur minimale de métallisation d'une ligne, c'est à dire le respect d'une densité surfacique minimale de cuivre sur la surface d'une couche de la plaque (wafer) au moment de son polissage.

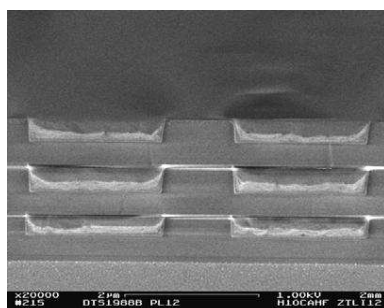


Figure 3-20 : Dummies insérés à chaque niveau de métallisation afin de respecter une densité de cuivre minimale sur une surface à polir

Pour respecter cette densité surfacique minimale de cuivre sur toute la plaque, on a recourt à l'insertion, dans le dessin du circuit, d'éléments métalliques de cuivre couramment nommés par le terme anglais « dummies » (Figure 3-20). Ils n'ont pas de rôle électrique, leur présence est justifiée par des aspects mécaniques afin d'assurer une planarisation excellente des couches lors du polissage par CMP.

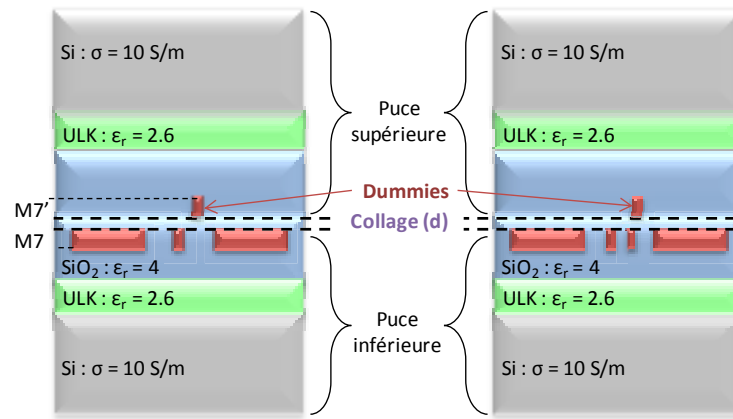


Figure 2-21 : Vue schématique d'un empilement Face to Face avec dummies. A gauche, le cas d'une interconnexion isolée, à droite, le cas de deux interconnexions couplées.

### II.3.1.b.1 Ligne de transmission isolée.

- Influence de la position d'un dummy vis-à-vis de celle de l'interconnexion sur le niveau M7 de la puce supérieure

Afin de déterminer l'impact sur le retard à 50% de la position, vis-à-vis de l'interconnexion CPW, d'un dummy situé au niveau M7 de la puce supérieure, plusieurs cas sont étudiés. On nomme  $n$  la distance horizontale séparant le dummy de l'interconnexion :

- Le dummy est situé juste au dessus de la ligne signal de la puce inférieure ( $n = 0$  nm).
- Le dummy est situé à une distance  $n$  de la ligne.
  - $n = 800$  nm : le dummy est situé entre la ligne signal et la masse.
  - $n = 2900$  nm : le dummy est situé au dessus de la masse.
- Sans dummy.

L'espacement  $S$  entre la ligne signal CPW et le ruban de masse reste constant durant l'étude et égal à 1200 nm.

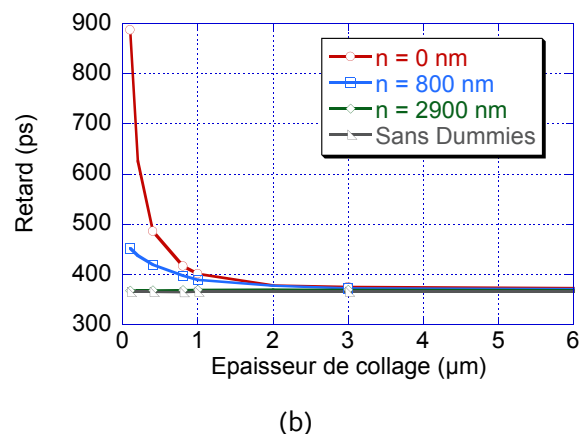
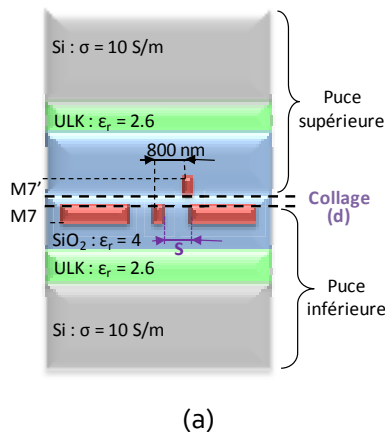


Figure 3-22 : Influence de la position d'un dummy sur le retard à 50%. (a) Vue schématique de l'interconnexion CPW en empilement Face to Face. (b) Retard en fonction de l'épaisseur de collage pour différentes positions du dummy.

D'après la Figure 3-22, on constate que pour de faibles épaisseurs de collage, plus le dummy est proche de la ligne signal, plus le retard est important. Inversement, lorsque le dummy est suffisamment éloigné de la ligne signal (ex :  $n = 2900$  nm), il n'a aucune influence sur le retard, quelque soit l'épaisseur de collage.

De plus, ce retard tend à se stabiliser, lorsque l'épaisseur de collage est supérieure à  $3\text{ }\mu\text{m}$ , à une valeur proche de celle d'une structure sans dummy. Ce qui indique qu'à partir d'une certaine épaisseur de collage, la présence de dummies n'a plus d'influence.

Pour illustrer ce résultat, le Tableau 3-5 indique les écarts relatifs sur le retard entre une structure d'interconnexion ne comprenant pas de dummies et une structure en comprenant, ceci pour trois positions de dummy différentes :  $n = 0$  nm,  $n = 800$  nm et  $n = 2900$  nm et pour trois épaisseurs de collage :  $d = 0,8\text{ }\mu\text{m}$ ,  $d = 3\text{ }\mu\text{m}$  et  $d = 6\text{ }\mu\text{m}$ .

Tableau 3-5 : Ecart relatif sur le retard à 50% entre différentes architectures d'interconnexions CPW.

	$d = 0,8\text{ }\mu\text{m}$	$d = 3\text{ }\mu\text{m}$	$d = 10\text{ }\mu\text{m}$
$n = 0$ nm	13,9 %	2,6 %	0,94 %
$n = 800$ nm	8,8 %	1,72 %	0,83 %
$n = 2900$ nm	0,92 %	0,99 %	0,67 %

On constate que lorsque la ligne signal est située juste au dessous ( $n=0$  nm) et à 800 nm d'un dummy l'écart relatif des retards diminue significativement lorsque l'on augmente l'épaisseur de collage. Lorsque le dummy est situé au-dessus de la masse, l'écart relatif ne dépasse pas 1%. Ces résultats indiquent plusieurs pistes pour minimiser le retard à 50%. Par exemple, imposer une zone d'exclusion de dummies au dessus la ligne signal lorsque l'épaisseur de collage est faible ( $< 1\text{ }\mu\text{m}$ )

Critère d'optimisation n°5: Eloigner les dummies de la ligne signal. Soit grâce à l'épaisseur de collage, soit en insérant une zone d'exclusion de dummies.

Par exemple, lorsqu'un dummy est situé juste au dessus de la ligne signal et que l'épaisseur de collage est de 100 nm, le retard est de l'ordre de 900 ps. Ce retard peut être réduit soit en éloignant le dummy, pour  $n = 800$  nm le retard diminue de 48 %, soit en augmentant l'épaisseur de collage, pour  $d = 1\text{ }\mu\text{m}$  le retard diminue de 55 %.

- Influence de l'espacement entre les conducteurs de l'interconnexion CPW pour un empilement Face to Face

Deux interconnexions CPW du BEOL présentant des espacements  $S$  différents sont étudiées lors d'un empilement Face to Face afin de déterminer le retard à 50%, en fonction de l'épaisseur de collage (Figure 3-21):

- Pour la première  $S = 800$  nm, Figure 3-23 (a).
- Pour la seconde  $S = 1200$  nm.

Cette étude permet de mettre en évidence une relation entre l'espacement des conducteurs et la valeur de l'épaisseur de collage,  $d_{\text{seuil}}$  à partir de laquelle le substrat de la puce supérieure n'a plus d'influence lors d'un empilement Face to Face.

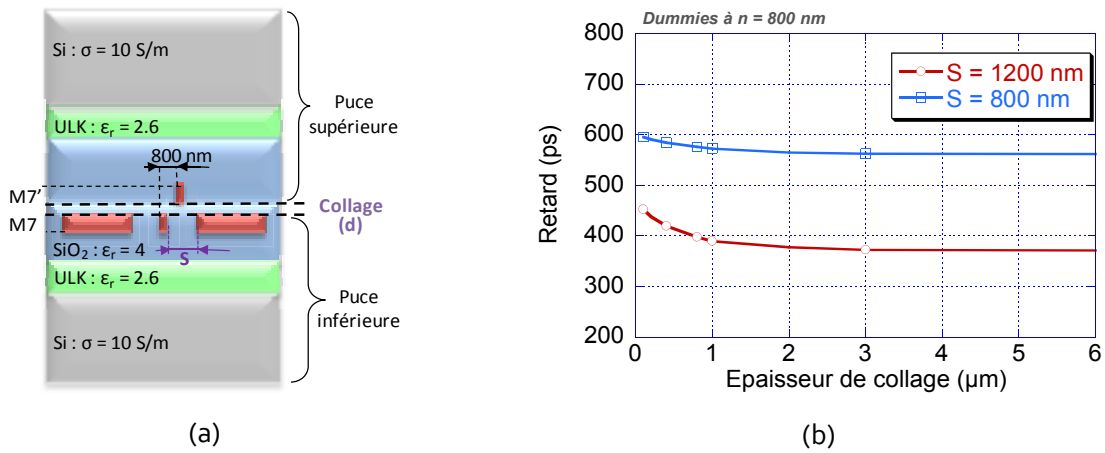


Figure 3-23 : Influence de l'espacement entre les conducteurs d'une interconnexion CPW du BEOL pour un empilement Face to Face de deux puces. (a) Vue schématique de l'interconnexion du BEOL. (b) Retard en fonction de l'épaisseur de collage pour différents espacements entre conducteurs.

Pour les mêmes raisons que dans le cas d'un collage Back to Face, plus l'espacement ligne-masse est grand, plus le retard à 50% est faible. Dans cette configuration précise, nous pouvons déterminer une relation empirique liant l'espacement entre les conducteurs ( $S$ ) et l'épaisseur de collage limite,  $d_{\text{seuil}}$  :

$$d_{\text{seuil}} = 3 \cdot S \quad \text{Équation 3-12}$$

En effet, lorsque  $S$  est égal à 800 nm,  $d_{\text{seuil}}$  est égale à 2,4  $\mu\text{m}$  et lorsque  $S$  est égal à 1200 nm,  $d_{\text{seuil}}$  est égale à 3,6  $\mu\text{m}$ .

Critère d'optimisation n°6 : Dans une configuration Face to Face, l'épaisseur de collage doit être trois fois plus élevée que l'espacement ligne-masse de l'interconnexion CPW du BEOL.

Par exemple, pour des métallisations espacées de 800 nm, le fait de passer d'une épaisseur de collage de 100 nm à 200 nm (zone où le substrat exerce une influence), permet de réduire le retard de 8 %. Et le fait de passer à un espacement de 1200 nm pour une épaisseur de collage de 200 nm, permet une diminution du retard de 25 %.

### II.3.1.b.2 Interconnexions couplées dans le BEOL pour un empilement Face to Face

Afin d'étudier l'influence d'un dummy situé au niveau M7 de la puce supérieure sur le retard de propagation des interconnexions couplées enfouies dans le BEOL de la puce inférieure, deux types d'interconnexions ont été simulés :

- Une interconnexion sans dummies.
- Une interconnexion avec dummies ( $n = 800$  nm).

L'espacement  $S$  entre la ligne signal et le ruban de masse reste constant et égal à  $1200$  nm.

Les figures 3-22 présentent les retards à 50 % et le niveau de diaphonie obtenus pour les différents types d'excitation des interconnexions couplées en fonction de l'épaisseur de collage.

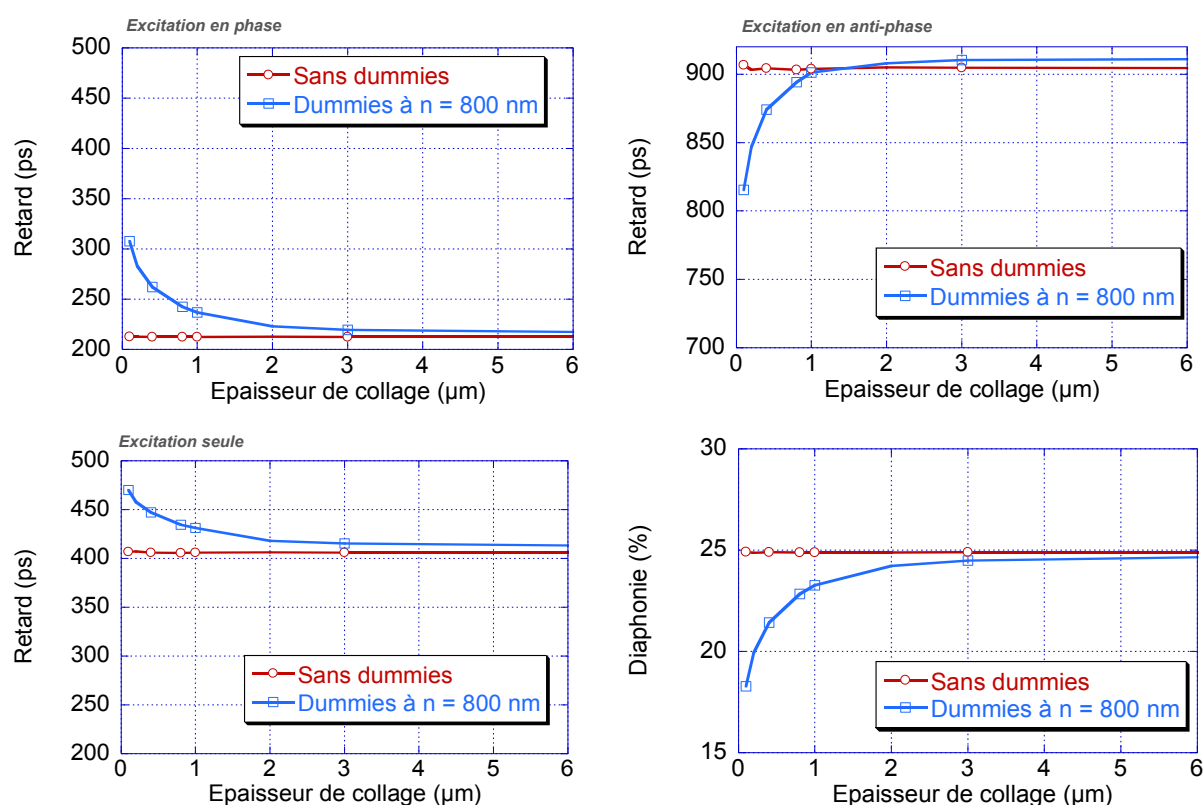


Figure 3-24 : Temps de retard pour chaque type d'excitation et diaphonie en fonction de l'épaisseur de collage ; Avec et sans dummies.

En configuration Face to Face, le retard et la diaphonie restent constants lorsqu'il n'y a pas de dummies sur la puce supérieure. La présence d'un dummy augmente le retard et diminue la

diaphonie, à mesure que l'épaisseur de collage diminue. Ce commentaire n'est valable que lors d'une excitation en phase et d'une excitation seule car, dans le cas d'une excitation en anti-phase, malgré des valeurs de retard très fortes, la présence d'un dummy diminue le retard lorsque l'épaisseur de collage diminue. Cette observation confirme le critère d'optimisation n°4.






### II.3.2 Synthèse des règles de dessin

Définir des règles d'optimisation génériques, pour toutes dimensions d'interconnexions, tout type de diélectrique, toutes formes et dimensions de dummies et pour chaque architecture d'empilement de puces (Face to Face ou Back to Face), n'est malheureusement pas possible. Néanmoins :

- nous pouvons énoncer plusieurs règles générales de design à respecter, c'est-à-dire les tendances à suivre pour minimiser les retards et la diaphonie et nous quantifions ci-après (Tableau 3-6) les gains obtenus sur ces deux critères de performance dans le cas des topologies d'interconnexions et types d'empilement que nous avons étudiés.

- ces études paramétriques nous montrent qu'il est possible de quantifier et optimiser les retards de propagation et diaphonie (et d'autres critères de performances comme par exemple le temps de montée 10%-90%) pour toute configuration d'interconnexion dès lors que les topologies, dimensions, architectures d'empilement sont données.

Tableau 3-6 : Synthèse des critères d'optimisations.

Contexte	Critère d'optimisation	Quantification par des exemples
<b>Interconnexion isolée</b> <b>Topologie CPW</b> <b>Report Back to Face</b>	 de l'épaisseur de collage	Passer d'une épaisseur de collage de 100 nm à une épaisseur de 200 nm, avec un espacement entre les métallisations de 800 nm.  =>  du retard de 17 %
	 de l'espacement entre les métallisations	Passer d'un espacement entre les métallisations de 800 nm à 1200 nm, pour une épaisseur de collage de 200 nm.  =>  du retard de 11 %
<b>Interconnexion isolée</b> <b>Report Back to Face</b>	Changement de topologie : Coplanaire : CPW Microruban : MSW	Pour une épaisseur de collage de 100 nm il faut privilégier une topologie MSW. =>  du retard de 34 %  Pour une épaisseur de collage de supérieure à 1



		<p><math>\mu\text{m}</math> il faut privilégier une topologie CPW.</p> <p>=&gt; ↘ du retard de 4 %</p>
<p><b>Interconnexion couplées</b></p> <p><b>Topologie CPW</b></p> <p><b>Report Back to Face</b></p>	<p>↗ de l'épaisseur de collage</p>	<p>Passer d'une épaisseur de collage de 100 nm à une épaisseur de 200 nm, avec un espacement entre les métallisations de 800 nm.</p> <p>=&gt; ↘ du retard de 25 % (lors d'une excitation seule).</p> <p>=&gt; ↗ la diaphonie de 3,5 %</p>
	<p>↗ de l'espacement entre les métallisations</p>	<p>Passer d'un espacement entre les métallisations de 800 nm à 1200 nm, pour une épaisseur de collage de 1 <math>\mu\text{m}</math>.</p> <p>=&gt; ↘ du retard de 12 % (lors d'une excitation seule).</p> <p>=&gt; ↗ la diaphonie de 2,2%</p>
<p><b>Interconnexion couplée</b></p> <p><b>Report Back to Face</b></p>	<p>Changement de topologie : Coplanaire : CPW Microruban : MSW</p>	<p>Pour une épaisseur de collage de 100 nm il faut privilégier une topologie MSW.</p> <p>=&gt; ↘ du retard de 23 %</p> <p>=&gt; ↗ la diaphonie de 4,5 %</p> <p>Pour une épaisseur de collage de supérieure à 800 nm il faut privilégier une topologie CPW. (pour une épaisseur de collage de 2 <math>\mu\text{m}</math>)</p> <p>=&gt; ↘ du retard de 14 %</p> <p>=&gt; ↗ la diaphonie de 2 %</p>
<p><b>Interconnexion isolée avec dummy</b></p> <p><b>Topologie CPW</b></p> <p><b>Report Face to Face</b></p>	<p>↗ de l'épaisseur de collage</p>	<p>Avec un dummy situé juste au dessus de l'interconnexion et pour une épaisseur de collage passant de 100 nm à 1 <math>\mu\text{m}</math>.</p> <p>=&gt; ↘ du retard de 55 %.</p>
	<p>↗ de l'espacement entre les métallisations</p>	<p>Avec un dummy distant de 800 nm de l'interconnexion. Passer d'un espacement entre les métallisations de 800 nm à 1200 nm, pour une épaisseur de collage de 100 nm.</p>

		=> ↘ du retard de 25 %.
	↗ de l'espacement entre l'interconnexion et le dummy	Avec une épaisseur de collage de 100 nm et pour un dummy qui passe du dessus de l'interconnexion à une distance de 800 nm.  => ↘ du retard de 48 %

### III Les interconnexions verticales.

Les interconnexions verticales regroupent les éléments intégrés permettant de véhiculer un signal d'une puce à une autre puce ou entre chaque face d'une puce. Dans le cas de communication intra-puce, on parlera de TSV alors que dans le cas de communication inter-puces, on parlera de Cu-Pillar. Dans cette partie les performances électriques de plusieurs TSV et plusieurs Cu-Pillars, de diverses densités d'intégration sont étudiées.

#### III.1 Etudes prédictives des performances des Cu-Pillars.

Ces interconnexions verticales servent de liens électriques dans le cas d'un empilement entre deux niveaux. Elles doivent être adaptées aux types de signaux à véhiculer entre deux puces ou entre une puce et un substrat afin de ne pas compromettre l'intégrité des signaux transmis.

##### III.1.1 Description des différents types de Cu-Pillar.

Dans le chapitre 1, deux technologies de réalisation de Cu-Pillar ont été définies.

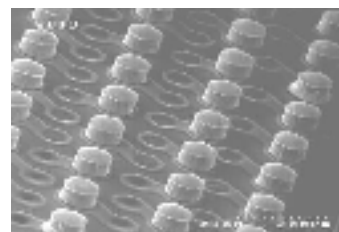
La première est utilisée pour la transmission des signaux entre deux puces. Les Cu-Pillar sont, soit constitués de trois couches de métallisations, deux en cuivres et une centrale en SnAg (Figure 3-25-a), soit réalisés par collage direct (Figure 3-25-b). La seconde est utilisée pour la transmission des signaux entre une puce et un substrat. Les Cu-Pillar sont cette fois-ci constitués de deux couches de métallisations, une en cuivre du côté de la puce et une en SnAg du côté du substrat (Figure 3-25-c).



(a)



(b)



(c)

Figure 3-25 : Coupe SEM de Cu-Pillar.

(a) Puce à puce. (b) Puce à puce par collage directe. (c) Puce à substrat.

Un schéma illustratif (Figure 3-26) ainsi qu'un tableau récapitulatif (Tableau 3-7) synthétisent les principales caractéristiques des Cu-Pillar étudiés. Ils sont classés de la plus faible densité d'intégration à la plus forte.

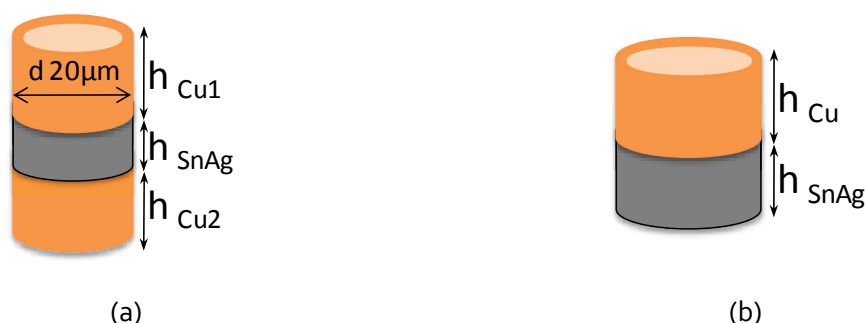


Figure 3-26 : Vue schématique d'un Copper-Pillar.  
(a) Communication puce à puce. (b) Communication puce à substrat.

On note  $h_{Cu1}$ ,  $h_{SnAg}$  et  $h_{Cu2}$ , les épaisseurs respectives de la première couche de cuivre, de la couche de SnAg et de la seconde couche de cuivre, le tout étant plongé dans un underfill de permittivité 3,3. L'espacement entre deux Cu-Pillar est noté  $S$ .

Tableau 3-7 : Synthèse des caractéristiques des Cu-Pillars

	<i>Cu-Pillar-d75</i>	<i>Cu-Pillar-d55</i>	<i>Cu-Pillar-d35</i>	<i>Cu-Pillar-d20</i>	<i>Cu-Pillar-d3</i>
<b>d (μm)</b>	75	55	35	20	3
<b>S (μm)</b>	180	150	100	40	10
<b><math>h_{Cu1}</math> (μm)</b>	45	45	17	17	/
<b><math>h_{SnAg}</math> (μm)</b>	35	35	15	15	/
<b><math>h_{Cu2}</math> (μm)</b>	0	0	0	10	/

Ils sont nommés en fonction de leur diamètre, par exemple un Cu-Pillar de diamètre 20 μm de hauteur totale 42 μm est noté Cu-Pillar-d20. Les trois premiers Cu-Pillar servent d'interface entre une puce et un substrat, un PCB (Printed Circuit Board) ou un BGA (Ball Grid Array). Le quatrième (Cu-Pillar-d20) est principalement utilisé pour faire l'interface entre deux puces. Le dernier Cu-Pillar-d3, est un cas particulier, il apparaît lors du collage direct de type Cu-Cu [Taibi, 2010].

### III.1.2 Analyse fréquentielle et modèles électriques des Cu-Pillars.

Lors des études menées, il apparaît que les Cu-Pillars ne sont pas le siège de phénomènes de propagation. Ceci est mis en évidence par l'évolution fréquentielle des différents termes de leur matrice ABCD, obtenue grâce à la matrice de diffusion  $[S]$  issue de simulations électromagnétiques.

Cette évolution fréquentielle des différents termes de la matrice ABCD des Cu-Pillars est donné en Figure 3-27.

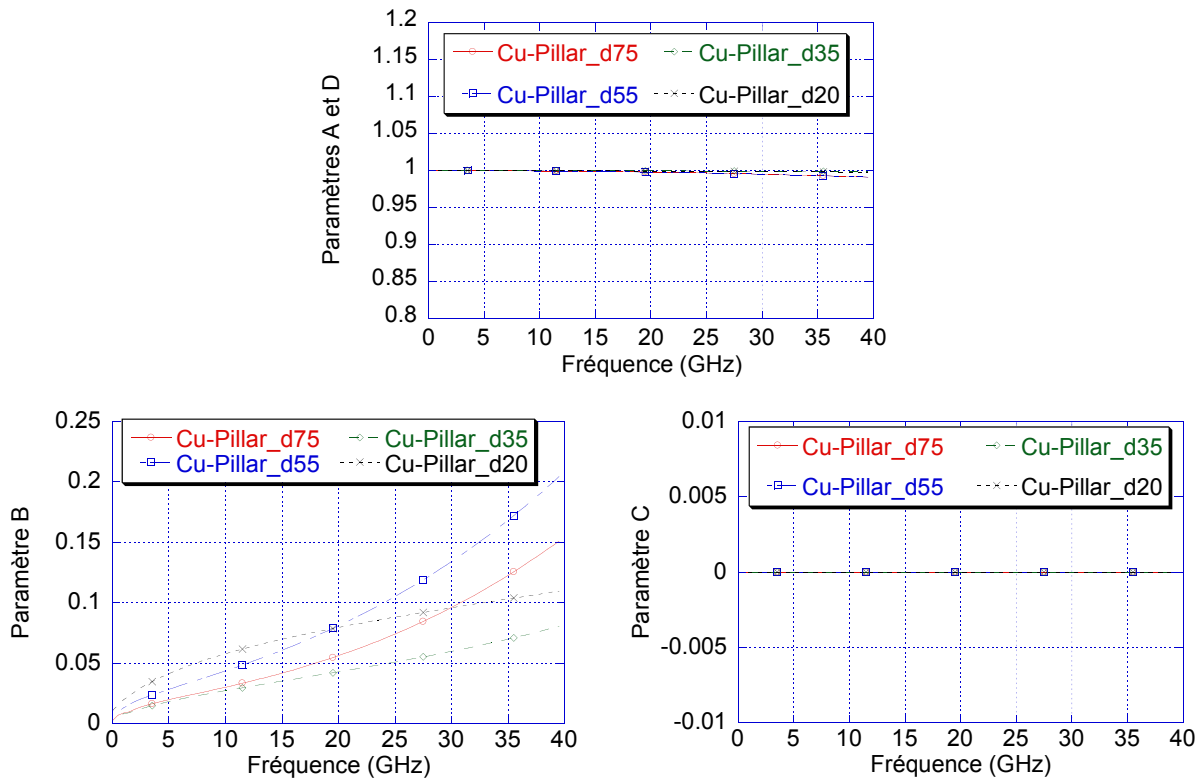


Figure 3-27 : Evolution fréquentielle des quatre termes de la matrice ABCD des Cu-Pillar.

Seul le paramètre B varie en fonction de la fréquence, les paramètres A et D sont égaux à 1 et le paramètre C est nul. Une telle évolution fréquentielle reflète qu'un Cu-Pillar se modélise par une impédance complexe et localisée [Pozar, 2005]. Dans notre cas, l'impédance est composée d'une résistance, équivalente à la somme des résistances statiques des métallisations et d'une capacité composée d'une capacité représentant le couplage entre deux Cu-Pillars et d'une capacité entre le Cu-pillar et les niveaux de métallisation situés, soit aux niveaux de la RDL pour les Cu-Pillar puce à substrat, soit au dernier niveau de métallisation de BEOL pour un Cu-Pillar puce à puce. La Figure 3-28 représente la manière de modéliser un Cu-Pillar.

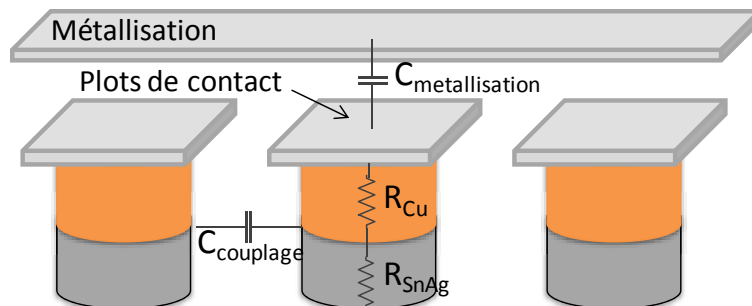


Figure 3-28 : Modélisation des éléments résistif et capacitif d'un Cu-Pillar

Dans la littérature de nombreuses études ont été menées afin d'obtenir les formules (Équation 3-13 et Équation 3-14) nécessaires à la détermination de ces deux valeurs [Sakurai, 1983].

$$R = R_{Cu} + R_{SnAg} = \frac{\rho_{Cu} C_{Cu} \cdot h_{Cu}}{\pi \cdot (d/2)^2} + \frac{\rho_{SnAg} C_{Cu} \cdot h_{SnAg}}{\pi \cdot (d/2)^2} \quad \text{Équation 3-13}$$

$$C = C_{couplage} + C_{métallisation} = \frac{2 \cdot \pi \cdot \epsilon_r \cdot \epsilon_0 \cdot h_{tot}}{\ln\left(\frac{S + d/2}{d/2}\right)} + \frac{\epsilon_r \cdot \epsilon_0 \cdot S_p}{e} \quad \text{Équation 3-14}$$

Le SiO<sub>2</sub> et l'underfill présentent respectivement une permittivité relative de 4,2 et 3,3.

S<sub>p</sub> correspond à la surface du plot de contact accueillant le Cu-Pillar.

e est l'espacement entre un Cu-Pillar et les métallisations situé au dessus.

Le Tableau 3-8, récapitule les valeurs de la résistance et de la capacité des différents Cu-Pillar.

Tableau 3-8 : Synthèse des valeurs caractéristiques des Cu-Pillar.

	<i>Cu-Pillar-d75</i>	<i>Cu-Pillar-d55</i>	<i>Cu-Pillar-d35</i>	<i>Cu-Pillar-d20</i>	<i>Cu-Pillar-d3</i>
<b>R (mΩ)</b>	1,15	2,1	2,2	7,5	0
<b>C<sub>couplage</sub> (fF)</b>	8,3	7,8	3	5	/
<b>C<sub>métallisation</sub> (fF)</b>	89	52	25	52	/
<b>C (fF)</b>	97	60	28	57	0

Les très faibles valeurs de R et C font que le retard engendré par les Cu-Pillar est généralement faible en comparaison de ceux induits par les autres briques d'interconnexions élémentaires et par les TSV en particulier. Néanmoins nous intégrerons ces modèles dans nos simulations de performances des chaînes complètes d'interconnexion lors d'un empilement 3D de puces afin de ne pas ignorer leurs effets dans le cas où ils sont associés à des interconnexions du BEOL ou de RDL très courtes.

### III.2 Etudes prédictives des performances des TSV

Dans le but de quantifier les performances électriques des différents types de TSV, cette partie est divisée en trois sous-thèmes. Dans un premier temps, le descriptif complet des TSV étudiés sera exposé, en termes de technologies 3D d'empilement et de leurs dimensions. Puis, à l'aide d'une étude fréquentielle, une comparaison de leurs paramètres R, L, C et G est effectuée. Pour finir, une

étude temporelle sera menée afin de déterminer le retard à 50 % introduit lors de la propagation d'un signal le long d'un TSV.

### III.2.1 Description des différents types de TSV.

Deux technologies sont couramment utilisées pour la réalisation d'un TSV. Avec la première, le TSV est fabriqué après le collage. Il est nommé « Via after bonding » [Druais, 2010] et n'est que partiellement rempli de métallisation (Figure 3-29-a). Dans notre étude il a un diamètre de 60  $\mu\text{m}$  et est nommé TSV-d60 [Guillon, 2009]. On classe aussi ces TSV dans la catégorie des TSV à faible ou moyenne densité. Dans la seconde technologie les TSV sont réalisés avant le collage, ils sont nommés « TSV Middle » et ils sont complètement remplis de cuivre (Figure 3-29-b). Nous étudions quatre de ces TSV de diamètres respectifs 10  $\mu\text{m}$ , 6  $\mu\text{m}$ , 5  $\mu\text{m}$  et 3  $\mu\text{m}$ . Ils font partie des TSV de moyenne à forte densité. Ils sont nommés TSV-d10, TSV-d6, TSV-d5 [Katti, 2011] et TSV-d3.

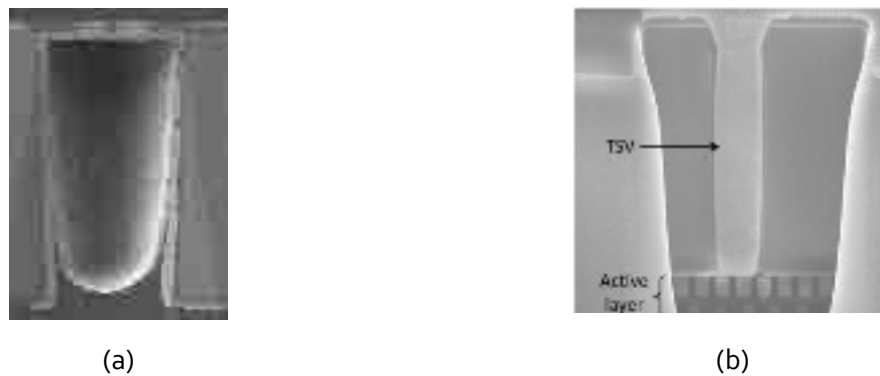


Figure 3-29 : Coupe SEM de TSV (a) Partiellement rempli. (b) Complètement rempli.

Cinq TSV de dimensions et de caractéristiques différentes sont étudiés. Un schéma illustratif donné en Figure 3-30 ainsi que le Tableau 3-7 synthétisent leurs principales caractéristiques. Ils sont classés de la plus faible densité d'intégration à la plus forte.

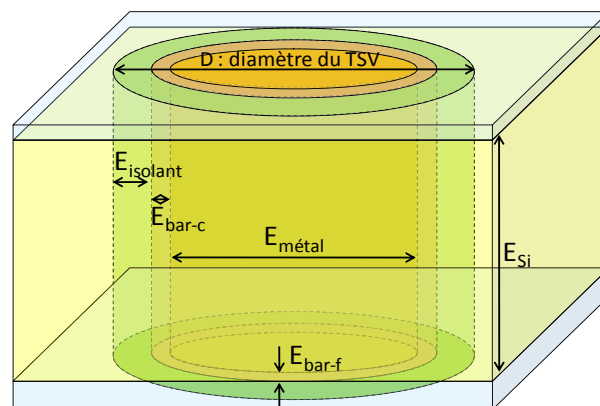


Figure 3-30 : Schéma illustratif d'un TSV.

On note  $E_{\text{métal}}$  l'épaisseur du métal remplissant partiellement ou complètement le TSV et  $\rho_{\text{métal}}$  sa résistivité.  $E_{\text{isolant}}$  est l'épaisseur de l'isolant situé sur les côtés du TSV, sa permittivité est notée  $\epsilon_{\text{isolant}}$ .  $E_{\text{bar-c}}$  et  $E_{\text{bar-f}}$  représentent les épaisseurs des barrières respectivement sur les flancs et en fond de via, où  $\rho_{\text{bar}}$  est la résistivité associée.

Tableau 3-9 : Synthèse des caractéristiques des TSV étudiés.

	<i>TSV-d60</i>	<i>TSV-d10</i>	<i>TSV-d6</i>	<i>TSV-d5</i>	<i>TSV-d3</i>
$D \text{ (}\mu\text{m)}$	60	10	6	5	3
$E_{\text{si}} \text{ (}\mu\text{m)}$	120	80	50	25	15
$E_{\text{isolant}} \text{ (nm)}$	500	300	150	150	200
$\epsilon_{\text{isolant}}$	5,2	5,2	5,2	5,2	5,2
$E_{\text{métal}} \text{ (}\mu\text{m)}$	4	10	6	5	3
$\rho_{\text{métal}} \text{ (}\mu\Omega\cdot\text{cm)}$	2	2	2	2	2
$E_{\text{bar-c}} \text{ (nm)}$	50	20	20	20	30
$E_{\text{bar-f}} \text{ (nm)}$	20	0	0	0	30
$\rho_{\text{bar}} \text{ (}\mu\Omega\cdot\text{cm)}$	1800	1800	1800	1800	1800

### III.2.2 Analyse fréquentielle et modèles des TSV

A partir des simulations électromagnétiques effectuées sur chaque TSV, leurs paramètres R, L, C et G donnés en Figure 3-31 ont pu être extraits en fonction de la fréquence.

Nous avons fait le choix de donner ces valeurs sous une forme « localisée », en effet les dimensions des TSV sont fixes, un diamètre correspond à une longueur et cette dernière n'a qu'une très faible marge de variation.

Cependant ce choix ne change rien au fait que de la propagation a lieu dans les TSV et qu'ils sont considérés de la même manière que des lignes de transmission.

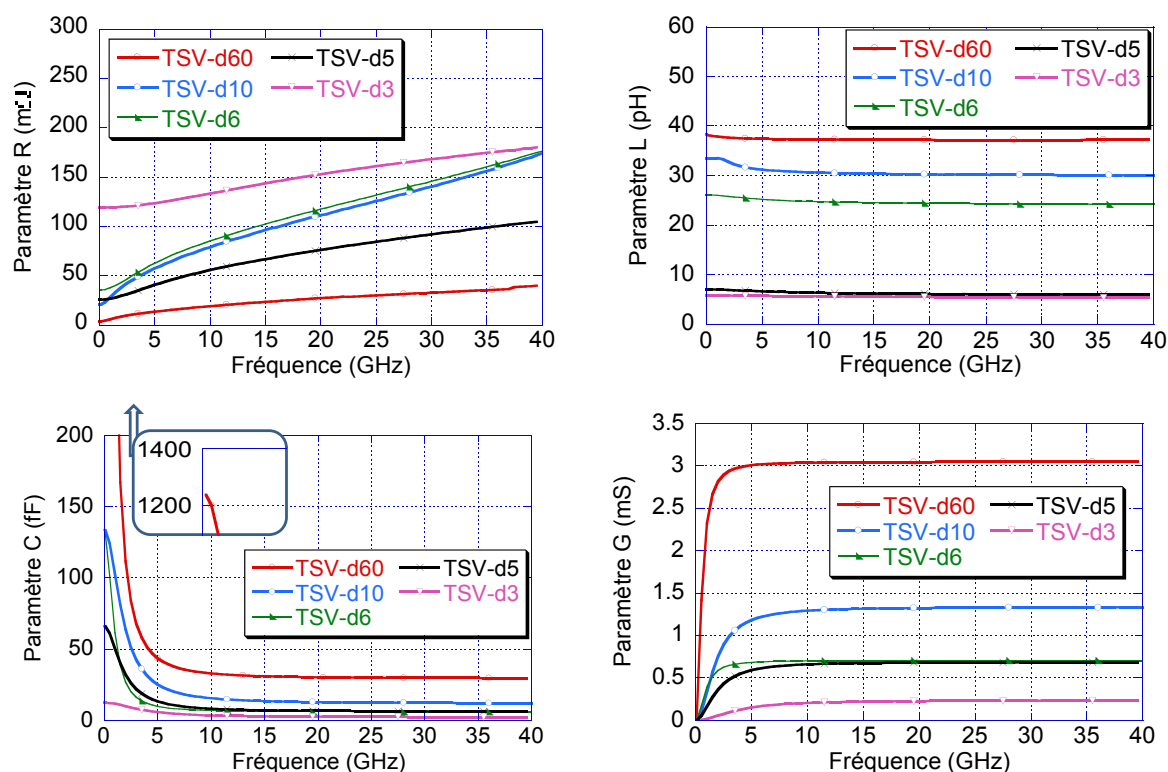


Figure 3-31 : Paramètres R, L, C et G des TSV

Le TSV de type « Via After Bonding », TSV-d60, est celui possédant les plus grandes dimensions. Par conséquent, sa résistance est naturellement inférieure à celle des autres TSVs sur toute la bande de fréquence. Inversement, le TSV le plus intégré (TSV-d3) a des dimensions très inférieures aux autres ce qui justifie sa valeur de résistance très élevée sur toute la bande de fréquence.

Remarque : lorsque le diamètre d'un TSV diminue, sa longueur diminue également (Tableau 3-9).

En ce qui concerne les TSV de diamètres 10, 6 et 5  $\mu\text{m}$ , leurs dimensions sont du même ordre de grandeur. Cependant le facteur de forme du TSV-d5 (égal à 5) est inférieur aux facteurs de forme des TSV-d6 et TSV-d10 (égaux à 8). Cet écart entre les facteurs de forme explique le fait que le TSV-d5 soit le moins résistif des trois. Toutefois, le TSV-d10 est légèrement plus large que le TSV-d6, cela confirme sa faible résistance.

Les valeurs caractéristiques de chaque type de TSV sont récapitulées dans le Tableau 3-10.

Tableau 3-10 : Valeurs caractéristiques des différents TSV.

	$R \text{ (m}\Omega\text{)}$	$L \text{ (pH)}$	$C_{ox} \text{ (fF)}$	$C_{si} \text{ (fF)}$	$G_{si} \text{ (mS)}$
TSV-d60	3,6	38,3	1237	30,3	3,1
TSV-d10	20,4	33,5	133,8	13,3	1,5
TSV-d6	35,6	26,0	131,7	6,3	0,7
TSV-d5	25,6	7,0	66,1	6,9	0,8
TSV-d3	119,1	5,7	12,4	3	0,3



### III.2.3 Analyse temporelle et performances de propagation des TSV.

Le retard engendré par chaque TSV est déterminé dans cette partie. Les drivers utilisés sont les mêmes que ceux utilisés dans l'étude des interconnexions horizontales.

La Figure 3-32, résume le retard à 50 % introduit sur un signal lors de sa propagation le long de chaque TSV, tous étant étudiés dans un environnement technologique identique.

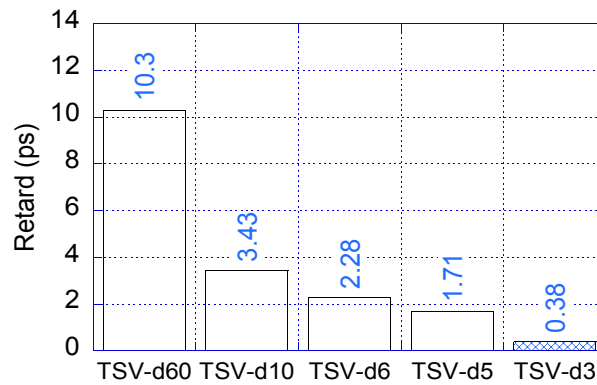


Figure 3-32 : Retard à 50 % engendré sur la propagation d'un signal par chaque TSV.

La Figure 3-32 montre que le retard diminue lorsque la densité d'intégration augmente. Contrairement aux interconnexions horizontales où la résistance est le paramètre prépondérant lorsque la densité d'intégration augmente, pour les interconnexions verticales la diminution des valeurs des paramètres parallèles ( $C_{ox}$ ,  $C_{si}$  et  $G_{si}$ ) est prépondérante par rapport à l'augmentation de la résistance.

## IV Inductances planaires

L'inductance est un composant passif fonctionnel directement intégré dans le BEOL des puces et son comportement électrique reste particulièrement sensible à son environnement proche. Cette étude met en évidence l'impact d'un empilement 3D sur les performances de ce composant passif. Les grandeurs étudiées pour caractériser l'inductance sont l'amplitude et la fréquence de résonance de son facteur de qualité. Les simulations ont été effectuées lors du report d'une seconde puce sur une puce de référence qui contient l'inductance (avant report), en architectures Back to Face et Face to Face.

Le facteur de qualité est défini par la relation suivante :

$$Q = -\frac{\text{Im}(Y_{11})}{\text{Re}(Y_{11})} \quad \text{Équation 3-15}$$

Comme illustré dans la Figure 3-33, représentant la topologie équivalente d'une inductance sur silicium, l'admittance  $Y_{11}$  d'une inductance, correspond à l'admittance obtenue lorsque l'on met en court-circuit un de ces deux accès.

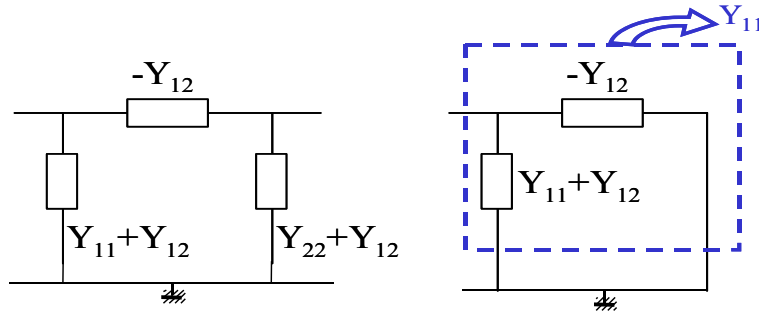


Figure 3-33 : Topologie équivalente d'une inductance sur silicium. (a) Vue comme un quadripôle. (b) Avec un accès en court circuit.

En pratique, les parties réelle et imaginaire du paramètre  $Y_{11}$ , sont déterminées à l'aide des paramètres S issus de simulations effectuées sous HFSS (Équation 3-16).

$$Y_{11} = \frac{1}{Z_0} * \frac{(1 - S_{11}) * (1 - S_{22}) - S_{12} * S_{21}}{(1 - S_{11}) * (1 - S_{22}) - S_{12} * S_{21}} \quad \text{Équation 3-16}$$

$Z_0$  étant l'impédance de normalisation (50 Ohm).

#### IV.1 Description de la structure de référence

La puce de référence qui contient l'inductance est constituée d'une inductance spirale 2 tours (Figure 3-34), dont l'espace inter-ruban  $S$  est de  $3 \mu\text{m}$ , le diamètre externe est de  $200 \mu\text{m}$ , la largeur des rubans est de  $20 \mu\text{m}$  et avec un plan de masse situé au niveau métallique M1.

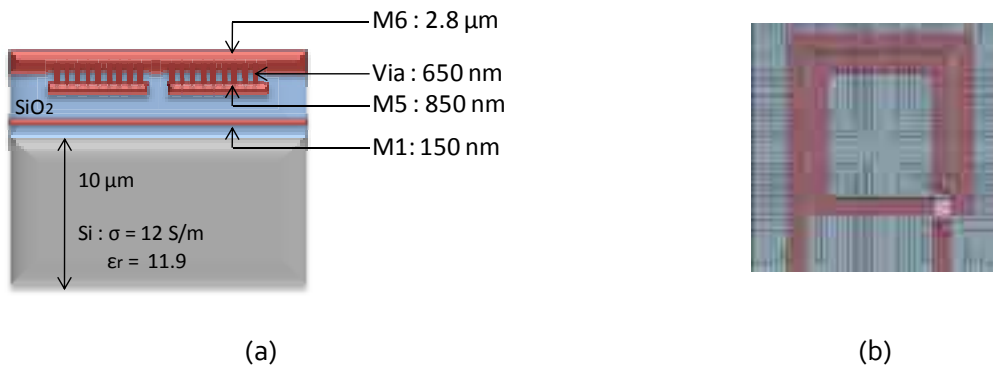


Figure 3-34 : Empilement technologique de la puce de référence intégrant l'inductance 2D planaire. (a) Vue en coupe. (b) Vue de dessus.

Comme indiqué sur la Figure 3-35, l'inductance sur la puce de référence présente un facteur de qualité d'amplitude maximale de 4,3 une fréquence de résonnance de 7 GHz et une bande passante de 12 GHz.

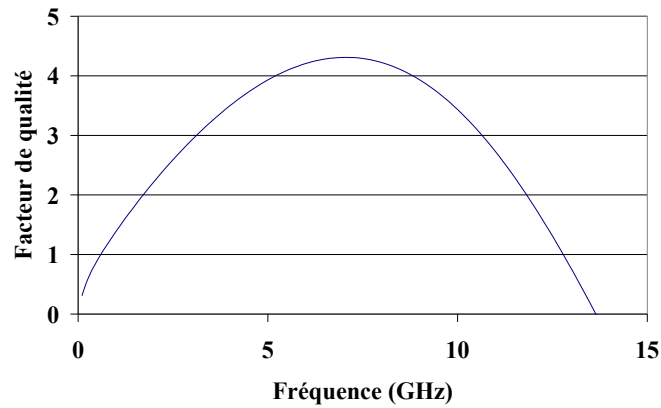


Figure 3-35 : Facteur de qualité de l'inductance située sur la puce de référence.

#### IV.1.1 Extraction des paramètres du modèle équivalent de l'inductance.

Le modèle de l'inductance présenté sur la Figure 3-36 est largement utilisé dans la littérature [Cao, 2003] [Murphy, 2005] et il met en évidence chaque élément caractéristique de l'inductance.

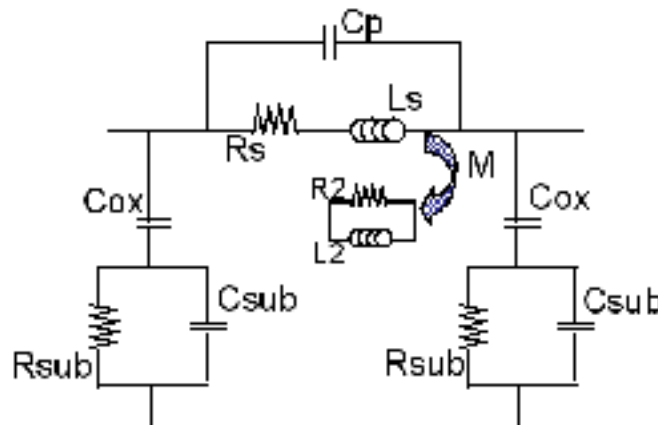


Figure 3-36 : Modèle électrique d'une inductance planaire intégrée sur silicium.

Dans ce modèle,  $L_s$  et  $R_s$  représentent respectivement l'inductance et la résistance série de la spirale. La capacité  $C_p$  traduit le couplage capacitif parasite entre les spires.  $C_{ox}$ ,  $C_{sub}$  et  $R_{sub}$  représentent les capacités issues des champs électriques régnant dans la couche d'oxyde et ceux pénétrant dans le substrat silicium. Une topologie de type transformateur, représentée par une mutuelle inductance ( $M$ ), une résistance  $R_2$  et une inductance  $L_2$ , permet d'introduire les pertes dues aux courants de Foucault dans le substrat silicium [Nguyen, 2008], [Tai, 2007], [Lee, 2006].

Il est possible de déterminer les différents éléments de ce modèle en faisant le parallèle avec les paramètres admittances, obtenus par les simulations HFSS, de la topologie équivalente représentée Figure 3-33.

#### IV.1.1.a Extraction des paramètres de la partie série

Suivant le modèle ci-dessus, l'impédance série  $-1/Y_{21}$  de la topologie équivalente à une inductance peut être approximée par la relation suivante :

$$Z_{série} = -1/Y_{12} = R_s + j\omega L_s + \frac{M^2 \omega^2}{R_2^2 + j\omega L_2^2} \quad \text{Équation 3-17}$$

Les paramètres résistif et inductif sont déterminés par les relations suivantes :

$$R = \text{real}(-1/Y_{12}) \quad \text{Équation 3-18}$$

$$L = -1/\omega \text{imag}(1/Y_{12}) \quad \text{Équation 3-19}$$

$R_s$  et  $L_s$  sont les valeurs en basse fréquence de  $R$  et  $L$ .

$M$ ,  $L_2$  et  $R_2$  sont remplacées par deux paramètres dont la signification physique est plus directe : le temps de relaxation  $\tau$  et le coefficient de couplage  $k$ . Ils sont définis par L. Nguyen [Nguyen, 2009]:

$$\begin{aligned} &\text{➤ } \tau : \text{ temps de relaxation.} \\ \tau &= L^2 / R^2 \end{aligned} \quad \text{Équation 3-20}$$

$$\begin{aligned} &\text{➤ } k : \text{ coefficient de couplage.} \\ k &= \frac{M^2}{L_s * L_2} \end{aligned} \quad \text{Équation 3-21}$$

En insérant ces relations dans l'Équation 3-17, on obtient :

$$Z_{série} = R + j.\omega.L = \left( R_s + \frac{k.L_s.\omega^2.\tau}{1 + \omega^2.\tau^2} \right) + j.\omega.\left( L_s - \frac{k.L_s.\omega^2.\tau^2}{1 + \omega^2.\tau^2} \right) \quad \text{Équation 3-22}$$

Par une identification de la partie imaginaire de l'Équation 3-22, nous pouvons remonter aux expressions permettant de déterminer les paramètres  $\tau$  et  $k$ .

$$K_{int} = 1 - L/L_s \quad \text{Équation 3-23}$$

$$\tau_{int} = \left[ \omega \sqrt{\frac{k L_s}{L_s - L}} - 1 \right]^{-1} \quad \text{Équation 3-24}$$

$k$  et  $\tau$  sont les valeurs extraites à partir des évolutions en fréquence de  $K_{int}$  et  $\tau_{int}$ . Lorsque celles-ci deviennent constantes.

#### IV.1.1.b Extraction des paramètres de la partie parallèle

Comme précédemment nous identifions la partie parallèle de la topologie équivalente (Figure 3-33 et Équation 3-25) à celle du modèle présenté Figure 3-36 (Équation 3-26). On obtient alors par identification les égalités présentées aux Équation 3-27 et Équation 3-28, qui nous permettent d'extraire les paramètres  $C_{ox}$ ,  $C_{si}$  et  $G_{si}$  (Équation 3-29 à Équation 3-31).

$$Y_{\text{équivalent}} = \text{real}(Y_{11} + Y_{12}) + j \cdot \text{imag}(Y_{11} + Y_{12}) \quad \text{Équation 3-25}$$

$$Y_{\text{modèle}} = \frac{\omega^2 \cdot G_{si}^2 \cdot C_{ox}^2}{G_{si}^2 + \omega^2 (C_{ox} + C_{si})^2} + j\omega \frac{C_{ox} G_{si} + \omega^2 C_{ox} C_{si} (C_{ox} + C_{si})}{G_{si}^2 + \omega^2 (C_{ox} + C_{si})^2} \quad \text{Équation 3-26}$$

$$G_{\text{equ}} = \text{real}(Y_{11} + Y_{12}) = \frac{\omega^2 \cdot G_{si}^2 \cdot C_{ox}^2}{G_{si}^2 + \omega^2 (C_{ox} + C_{si})^2} \quad \text{Équation 3-27}$$

$$C_{\text{equ}} = \text{imag}(Y_{11} + Y_{12}) \cdot \omega = \frac{C_{ox} G_{si} + \omega^2 C_{ox} C_{si} (C_{ox} + C_{si})}{G_{si}^2 + \omega^2 (C_{ox} + C_{si})^2} \quad \text{Équation 3-28}$$

$$\text{A basse fréquence } (\omega \rightarrow 0); \quad C_{ox} = C_{\text{equ}} \quad \text{Équation 3-29}$$

$$\text{A haute fréquence } (\rightarrow \infty); \quad C_{si} = \frac{C_{ox} C_{\text{equ}}}{C_{ox} - C_{\text{equ}}} \quad \text{Équation 3-30}$$

$$\text{A haute fréquence } (\rightarrow \infty); \quad G_{si} = \frac{G_{\text{equ}} (C_{ox} + C_{si})^2}{C_{ox}^2} \quad \text{Équation 3-31}$$

Le tableau ci-dessous synthétise les paramètres de l'inductance placée sur la puce de référence.

Tableau 3-11 : Valeurs caractéristiques de l'inductance placée sur la puce de référence.

$\tau$ (ps)	k	Rs (Ω)	Ls (nH)	Cp (fF)	Cox (fF)	Csub (fF)	Rsub (Ω)
250	0.45	3.7	1.8	0.2	138	500	4

#### IV.1.1.c Validation de la procédure d'extraction

En faisant la démarche inverse, c'est à dire en réinjectant les éléments déterminés précédemment, dans le modèle Figure 3-36, il est alors possible de calculer les admittances des parties série et parallèle et de les comparer aux admittances obtenues par simulation grâce aux paramètres [S].

La Figure 3-37, compare les résultats obtenus directement par la simulation sous HFSS à ceux obtenus en réinjectant dans le modèle les éléments déterminés. La première courbe correspond aux parties réelle et imaginaire de  $Y_{11}$ , éléments nous permettant de déterminer le facteur de qualité. La seconde, aux parties réelle et imaginaire de  $Y_{12}$ , éléments nous permettant de déterminer les paramètres résistif et inductif de l'inductance. Enfin, la dernière courbe correspond au facteur de qualité.

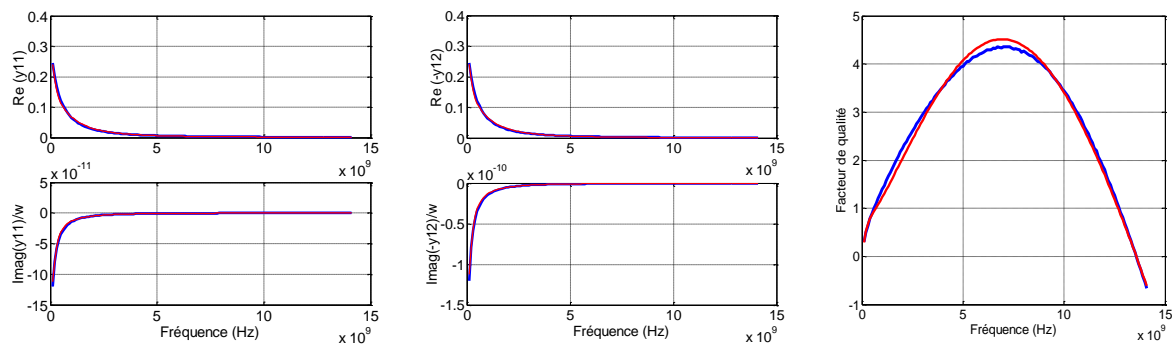


Figure 3-37 : Comparaison simulation et modèle (a)Paramètre  $Y_{11}$ . (b)Paramètre  $-Y_{12}$ . (c) Facteur de qualité.

On observe une très bonne concordance entre les résultats issus du modèle et ceux issus de la simulation sur les trois graphiques. Ce constat montre que les quelques approximations effectuées lors de l'extraction des paramètres du modèle n'ont pas d'influence majeure sur la pertinence des résultats.

## IV.2 Impact d'un empilement Face to Back sur les performances de l'inductance.

L'architecture Face to Back est réalisée en reportant un substrat de silicium sur la puce de référence contenant l'inductance avec un collage moléculaire  $\text{SiO}_2$ , d'épaisseur notée  $d$ .

### IV.2.1 Influence de l'épaisseur du substrat de silicium de la puce rapportée.

L'étude porte sur l'évolution du facteur de qualité de l'inductance en fonction de l'épaisseur du substrat de silicium de la puce rapportée en architecture Face to Back.

#### IV.2.1.a Description de la structure

L'empilement technologique de la puce de référence n'est bien entendu pas modifié. Seule l'épaisseur du substrat de silicium rapporté sur cette puce varie, de 5 à 100  $\mu\text{m}$ , sa conductivité étant égale à 12 S/m. Le collage entre les deux puces a été réalisé avec du  $\text{SiO}_2$  ( $\epsilon_r = 4$ ), l'étude est menée avec une épaisseur de collage  $d = 2\mu\text{m}$  (Figure 3-38). Le report est de type Face to Back

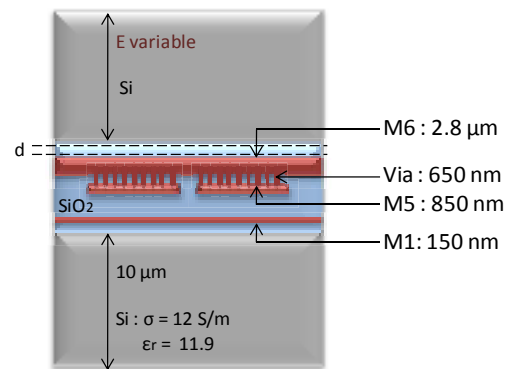


Figure 3-38 : Schéma en coupe des puces empilées en architecture Face to Back avec une épaisseur de silicium E variable pour le substrat rapporté sur l'inductance.

#### IV.2.1.b Résultats

Le premier constat visible en figure 3-37 est que le report d'une puce sur la puce de référence dégrade le facteur de qualité de l'inductance ; le facteur de qualité de l'inductance sur la puce de référence seule est toujours supérieur à celui obtenu dès le report d'une autre puce.

Comme on peut le remarquer sur la Figure 3-39, plus l'épaisseur E du substrat de silicium rapporté augmente, plus le facteur de qualité de l'inductance est dégradé. En effet lorsque l'épaisseur du silicium est faible, les lignes de champ se propagent dans le silicium puis dans l'air, alors qu'avec une épaisseur de substrat rapporté élevée, toutes les lignes de champs sont confinées dans le silicium. Comme le silicium provoque plus de pertes que l'air, on vérifie bien que le facteur de qualité décroît lorsque l'épaisseur du silicium augmente.

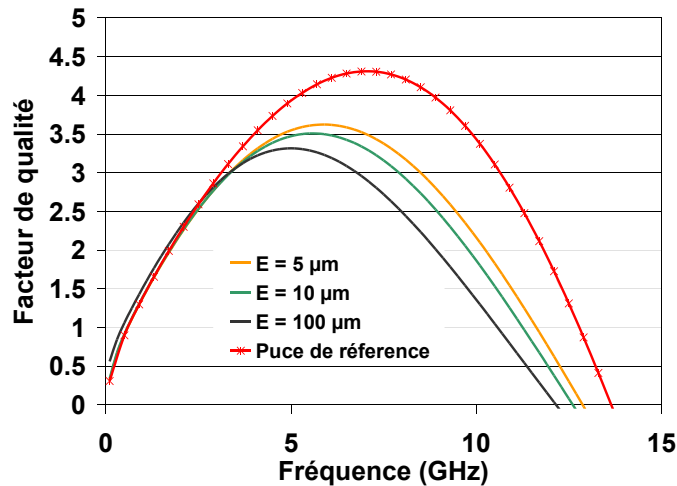


Figure 3-39 : Facteur de qualité en fonction de l'épaisseur du silicium de la puce rapportée

Lorsque E varie de 5 à 100 µm l'amplitude du facteur de qualité varie de 3,6 à 3,3, soit un écart relatif de 8,5 %. Concernant la bande passante, elle diminue de 11,2 GHz à 10 GHz, soit un écart relatif de 10,7 %.

Lorsque E est égal à 10 µm l'écart relatif avec la puce de référence est de l'ordre de 18,6 % pour l'amplitude et de 10,8 % pour la bande passante.

L'ajout d'un substrat de silicium a donc un impact majeur sur les performances des inductances. Il est important qu'il soit le plus mince possible.

Critère d'optimisation n°1 : Avoir un substrat de silicium de la puce rapportée le plus mince possible pour maintenir un facteur de qualité élevé de l'inductance.

Par exemple, amincir le substrat de la puce rapporté de 100 µm à 5 µm permet d'augmenter l'amplitude de facteur de qualité de 8,5 %.

## IV.2.2 Influence de la conductivité du silicium de la puce rapportée.

L'étude porte sur l'évolution du facteur de qualité de l'inductance en fonction de la conductivité du silicium de la puce rapportée.

### IV.2.2.a Description de la structure

L'empilement technologique de la puce de référence n'est bien entendu pas modifié. Nous n'avons fait varier que la conductivité du silicium du substrat de la puce rapportée, tout en gardant constant son épaisseur (10 µm). Le collage entre les deux puces a été réalisé avec du SiO<sub>2</sub> (κ = 4) et avec une épaisseur « d » de 2 µm (Figure 3-40). Le report est de type Face to Back.



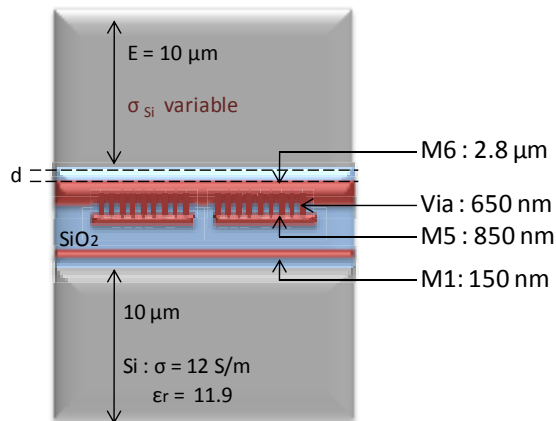


Figure 3-40 : Schéma en coupe du dispositif en Face to Back avec  $\sigma_{\text{Si}}$  variable.

#### IV.2.2.b Résultats

Lorsque la conductivité ( $\sigma_{\text{Si}}$ ) du silicium du substrat de la puce rapportée augmente, les pertes dans ce substrat augmentent et cela se traduit par une forte diminution du facteur de qualité de l'inductance (Figure 3-41).

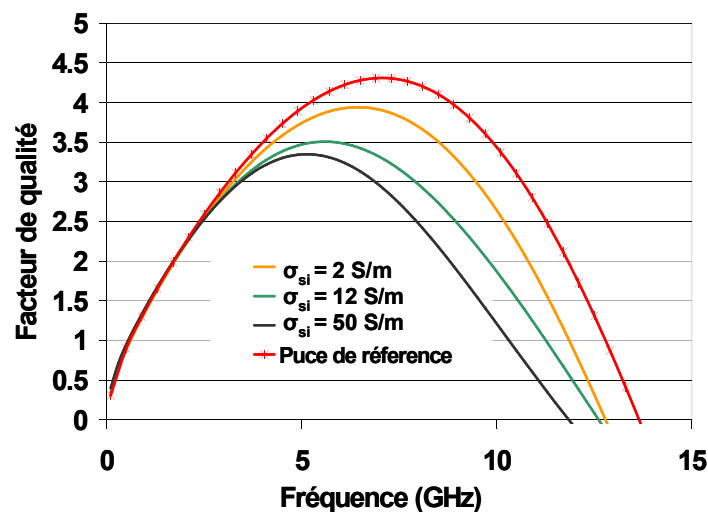


Figure 3-41 : Facteur de qualité en fonction de la conductivité du silicium de la puce rapportée

Lorsque la conductivité du silicium varie de 2 à 50 S/m on observe une diminution de l'amplitude du facteur de qualité de 3,9 à 3,3, soit un écart relatif de 15,2 %. Concernant la bande passante, on observe également une diminution puisqu'elle varie de 11,4 GHz à 9,8 GHz, soit un écart relatif de 14 %.

Lorsque  $\sigma_{\text{Si}}$  est égale 2 S/m l'écart relatif avec la puce de référence est de l'ordre de 8,8% pour l'amplitude et de 5 % pour la bande passante.

Critère d'optimisation n°2 : Avoir un substrat de silicium de la puce rapportée le moins conducteur possible pour maintenir un facteur de qualité élevé de l'inductance.

Par exemple, opter pour un substrat ayant une conductivité de 2 S/m plutôt qu'un substrat ayant une conductivité de 50 S/m permet d'augmenter l'amplitude de facteur de qualité de 15 %.

### IV.2.3 Influence de l'épaisseur de collage entre les deux puces empilées.

Nous étudions ici l'évolution du facteur de qualité de l'inductance en fonction de l'épaisseur  $d$  de collage entre les deux puces.

#### IV.2.3.a Description de la structure

L'empilement technologique de la puce de référence est conservé (figure 1.a). Seule l'épaisseur  $d$  de collage entre les deux puces, réalisé avec du  $\text{SiO}_2$  ( $\epsilon_r = 4$ ), sera variable. L'épaisseur de substrat rapporté (10  $\mu\text{m}$ ) et sa conductivité (12 S/m) restent constants (Figure 3-42). L'empilement est toujours de type Face to Back.

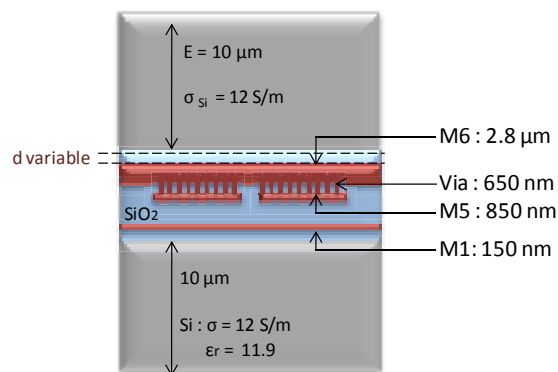


Figure 3-42 : Schéma en coupe de l'empilement en Face to Back avec  $d$  variable.

#### IV.2.3.b Résultats

Comme on peut le remarquer sur la Figure 3-43, plus l'épaisseur de collage ( $d$ ) augmente, plus le facteur de qualité de l'inductance s'améliore. En effet éloigner le silicium rapporté de l'inductance permet de diminuer les pertes dans le silicium rapporté.

Remarque : Cette épaisseur de collage symbolise également le diélectrique utilisé en face arrière accueillant les interconnexions de redistribution.

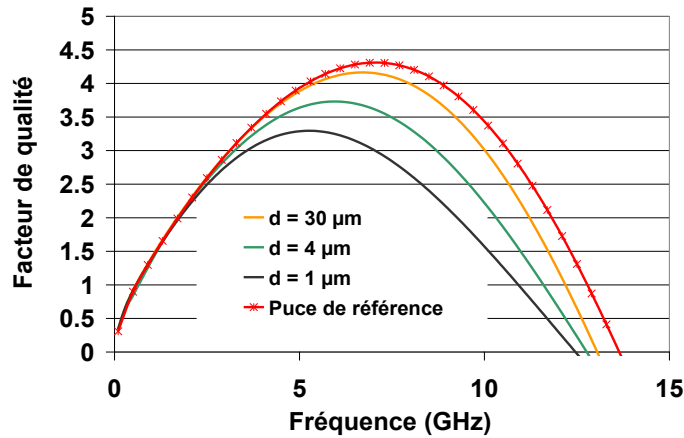


Figure 3-43 : Facteur de qualité en fonction de l'épaisseur de collage.

Lorsque  $d$  varie de 1 à  $30\mu\text{m}$ , on observe une augmentation de l'amplitude du facteur de qualité de 3,3 à 4,6, soit un écart relatif de 19,7 %. Concernant la bande passante, on observe également une augmentation puisqu'elle varie de 10,3 GHz à 11,6 GHz, soit un écart relatif de 12,2 %. Ainsi on peut constater que plus l'épaisseur de collage est élevée, plus l'inductance aura des performances se rapprochant de celle de la puce de référence sans pour autant les atteindre. En effet lorsque  $d = 30\mu\text{m}$  l'écart relatif avec la puce de référence est faible, de l'ordre de 4,6 % pour l'amplitude du facteur de qualité et de 3,3 % pour la bande passante.

Remarque : Les très grandes épaisseurs de collage prises pour cette étude ne sont pas réalisables mais nous permettent de dégager une forte tendance dans nos conclusions.

Critère d'optimisation n°3 : Une forte épaisseur de collage entre les puces empilées améliore le facteur de qualité de l'inductance.

Par exemple, une épaisseur de diélectrique de  $4\mu\text{m}$  entre l'inductance et le substrat de la puce supérieure, plutôt que de  $1\mu\text{m}$ , permet d'augmenter l'amplitude de facteur de qualité de 13 %.

### IV.3 Architecture Face to Face.

L'architecture Face to Face est réalisée en rapportant sur la puce de référence une puce dont le BEOL fait face à celui de la puce de référence, ce BEOL est constitué d'une couche diélectrique de  $\text{SiO}_2$  d'épaisseur  $6,2\mu\text{m}$  et d'un plan métallique  $\text{M1}'$ . Le collage est de type moléculaire  $\text{SiO}_2$ .

#### IV.3.1 Influence de l'épaisseur et de la conductivité du substrat de silicium de la puce rapportée.

L'étude porte sur l'évolution du facteur de qualité en fonction de l'épaisseur et de la conductivité du silicium de la puce rapportée (Figure 3-44).

#### IV.3.1.a Description de la structure

L'empilement technologique de la puce de référence est resté le même. Nous n'avons fait varier dans un premier temps que l'épaisseur du silicium de la puce rapportée, tout en gardant constant sa conductivité (12 S/m) et dans un second temps nous n'avons fait varier que la conductivité du silicium de la puce rapportée, tout en gardant constant son épaisseur (10  $\mu\text{m}$ ). Ces études ont été réalisées pour une épaisseur de collage  $d=1\mu\text{m}$ . Le collage entre les deux puces a été réalisé avec du  $\text{SiO}_2$  ( $\epsilon_r = 4$ ). La figure 3-42 présente l'empilement.

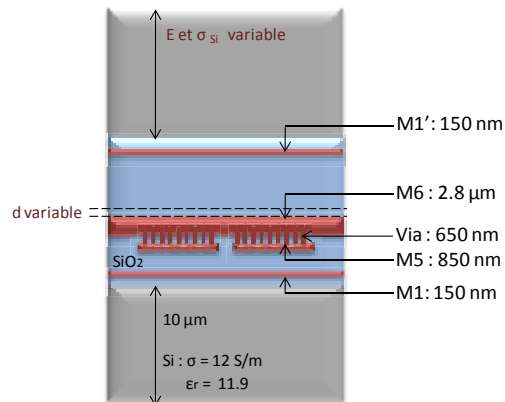


Figure 3-44 : Schéma en coupe du dispositif en face to face.

#### IV.3.1.b Résultat

Avec des variations de conductivité et d'épaisseur du silicium rapporté, on constate sur la figure 3-43 que le facteur de qualité et la bande passante restent constants quelques soient les évolutions des caractéristiques du substrat de la puce rapportée.

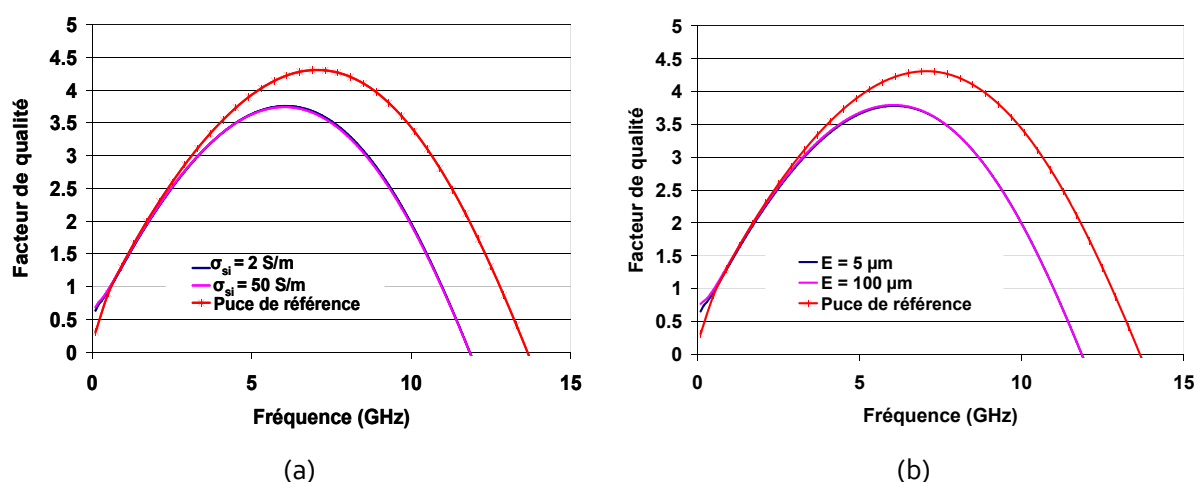


Figure 3-45 : Facteur de qualité en configuration Face to Face.

(a) Pour différentes conductivités de silicium. (b) Pour différentes épaisseurs de silicium.

En effet dans cette configuration les plans de masse au dessus et en dessous de l'inductance masquent les substrats de silicium et suppriment les courants de Foucault. Les champs sont donc confinés dans le  $\text{SiO}_2$ . Le substrat de silicium n'a plus d'influence sur les performances de l'inductance.

## V Conclusion

Dans ce chapitre, toutes les briques élémentaires d'interconnexion pouvant constituer une chaîne d'interconnexion en intégration 3D ont été étudiées. De vastes études paramétriques visant à analyser ces briques élémentaires (interconnexions de RDL, TSV, interconnexions de BEOL, Cu-Pillar et inductance) en fonction de leur design et des caractéristiques des matériaux présents dans les empilements ont permis de quantifier les valeurs de composants présents dans leurs modèles électriques équivalents et de prédire leurs performances en termes de retard, niveau de diaphonie ou de facteur de qualité.

Dans le cas des interconnexions de RDL, la performance des interconnexions à haute densité, intégrée dans du BCB a été démontrée. Pour les interconnexions du BEOL, le retard engendré dépend principalement de la longueur de l'interconnexion. Elle doit être inférieure à 1,4 mm afin de concurrencer les performances des interconnexions de RDL. Le retard induit par chaque TSV dépend de sa densité d'intégration : Un TSV intégré à haute densité introduit beaucoup moins de retard qu'un TSV intégré à faible densité. Concernant les Cu-Pillars, il a été observé que le retard induit est peu pénalisant par rapport au retard engendré par les autres briques.

Un modèle complet d'une inductance planaire 2 ports a également été exposé afin d'évaluer les performances des inductances insérées dans une architecture 3D.

De plus, deux études paramétriques poussées, sur l'impact d'une architecture 3D sur les performances de composants déjà présents en intégration 2D (les interconnexions horizontales et les inductances), ont permis de définir plusieurs règles d'optimisation lors d'une intégration 3D :

Pour les interconnexions horizontales :

- 1) Dans une configuration Face to Back, l'épaisseur de collage doit être deux fois plus élevée que l'espacement ligne-masse pour une interconnexion de type CPW. Dans le cas contraire elle doit être maximisée.

Passer d'une épaisseur de collage de 100 nm à une épaisseur de 200 nm, avec un espacement entre les métallisations de 800 nm permet de réduire le retard de 17 %.

- 2) L'espacement ligne-masse des interconnexions CPW doit être maximisé.

Passer d'un espacement entre les métallisations de 800 nm à 1200 nm, pour une épaisseur de collage de 200 nm permet de réduire le retard de 11 %

- 3) L'architecture des interconnexions doit être adaptée aux contraintes de fabrication.

Pour une épaisseur de collage de 100 nm, privilégier une topologie MSW permet de réduire le retard de 34 %

- 4) Un compromis doit être trouvé entre l'optimisation du retard et de la diaphonie.

Passer d'une épaisseur de collage de 100 nm à une épaisseur de 200 nm, avec un espacement entre les métallisations de 800 nm permet de réduire le retard de 25 % (lors d'une excitation seule) mais augmente le niveau de diaphonie de 3,5 %.

Passer d'un espacement entre les métallisations de 800 nm à 1200 nm, pour une épaisseur de collage de 1  $\mu$ m permet de réduire le retard de 12 % (lors d'une excitation seule) mais augmente le niveau de diaphonie de 2,2%.

- 5) Les dummies doivent être éloignés de la ligne signal, soit grâce à l'épaisseur de collage, soit par une zone d'exclusion.

Avec une épaisseur de collage de 100 nm et pour un dummy qui passe du dessus de l'interconnexion à une distance de 800 nm, le retard diminue de 48 %.

- 6) Dans une configuration Face to Face, l'épaisseur de collage doit être trois fois plus élevée que l'espacement ligne-masse. Dans le cas contraire elle doit être maximisée.

Avec un dummy situé juste au dessus de l'interconnexion et pour une épaisseur de collage passant de 100 nm à 1  $\mu$ m, le retard diminue de 55 %.

Pour les inductances, en configuration Face to Back:

- 1) Le substrat de silicium rapporté doit être le plus mince possible.

Amincir le substrat de la puce rapporté de 100  $\mu$ m à 5  $\mu$ m permet d'augmenter l'amplitude de facteur de qualité de 8,5 %.

- 2) La conductivité du silicium doit être la plus faible possible.

Opter pour un substrat ayant une conductivité de 2 S/m plutôt qu'un substrat ayant une conductivité de 50 S/m permet d'augmenter l'amplitude de facteur de qualité de 15 %.

- 3) Il faut privilégier les fortes épaisseurs de collage.

Par exemple, une épaisseur de diélectrique entre l'inductance et le substrat de la puce supérieure, de 4  $\mu$ m plutôt que de 1  $\mu$ m, permet d'augmenter l'amplitude de facteur de qualité de 13 %.

# Chapitre 4

## Intégrité de transmission des signaux sur une chaîne complète d'interconnexions dans les empilements 3D de circuits intégrés

---

Ce court chapitre aborde la thématique de l'intégrité des signaux transmis à travers une chaîne d'interconnexions dans un empilement 3D de puces.

La première partie donne accès aux méthodologies utilisées pour associer les différentes briques élémentaires d'interconnexion (RDL, BEOL, TSV, Cu-Pillar) afin de constituer des chaînes complètes d'interconnexions représentatives de celles rencontrées dans les circuits 3D, selon les différentes alternatives possibles d'empilements. Les techniques de simulations permettant de passer du domaine fréquentiel au domaine temporel et de simuler les performances de ces chaînes d'interconnexions, sont ensuite exposées.

Les deuxième, troisième et quatrième parties, mettent en scène différent scénario, afin de déterminer les grands axes à privilégier lors d'une intégration 3D.

- La deuxième partie est intitulée stratégie de routage. Elle met en évidence le meilleur chemin à emprunter pour transmettre les signaux au travers d'une puce.
- La troisième partie est intitulée stratégie d'orientation. Elle détermine quelle est la meilleure orientation pour la puce centrale, lors de l'empilement de trois puces.
- La dernière partie intitulée stratégie technologique, évalue les éléments d'interconnexions les plus critiques dans une chaîne d'interconnexion 3D en fonction des densités d'intégration utilisables.

## Sommaire

<b>I</b>	<b>INTRODUCTION.....</b>	<b>139</b>
<b>II</b>	<b>CONVERSION DU DOMAINE FREQUENTIEL AU DOMAINE TEMPOREL .....</b>	<b>140</b>
<b>III</b>	<b>STRATEGIE DE ROUTAGE .....</b>	<b>141</b>
III.1	DESCRIPTION DES SCENARIOS DE ROUTAGE .....	142
III.2	RESULTATS.....	143
<b>IV</b>	<b>STRATEGIE D'ORIENTATION .....</b>	<b>144</b>
IV.1	DESCRIPTION DES SCENARIOS D'ORIENTATION DES PUCES .....	144
IV.2	RESULTATS.....	145
<b>V</b>	<b>STRATEGIE TECHNOLOGIQUE. ....</b>	<b>146</b>
V.1	DESCRIPTION DES SCENARIOS TECHNOLOGIQUES.....	147
V.2	RESULTATS.....	147
V.2.1	<i>Performance en termes de retard. ....</i>	<i>147</i>
V.2.2	<i>Influence relative de chaque brique élémentaire d'interconnexion dans une chaîne complète d'interconnexion.....</i>	<i>148</i>
<b>VI</b>	<b>CONCLUSION .....</b>	<b>150</b>



## I Introduction

L'un des enjeux de l'intégration 3D consiste à connaître, avec la plus grande exactitude possible, le comportement en termes d'intégrité dans la transmission des signaux des chaînes d'interconnexions reliant les puces entre elles et vers l'extérieur (substrat BGA). Une bonne connaissance des performances de l'ensemble des briques élémentaires constituant ces chaînes d'interconnexions, permet l'exploration d'une multitude de scénarios technologiques. En effet, en combinant les matrices ABCD de chaque brique élémentaire d'interconnexion, diverses chaînes d'interconnexion 3D peuvent être construites afin d'évaluer leurs performances.

Les applications visées par l'intégration 3D n'étant pas encore toutes définies, il est nécessaire de dessiner de grands axes, en bannissant certaines voies et en privilégiant des designs particuliers. Ce chapitre aborde l'intégration 3D au travers de plusieurs scénarios dans le but d'établir trois stratégies majeures : une stratégie pour le routage, une pour l'orientation et une liée à la technologie.

Le critère de performances choisi dans les études traitées dans ce chapitre est le retard à 50% mais notre méthodologie peut être appliquée à d'autres critères de performance, comme les temps de montée des signaux, les niveaux de signal transmis, etc.

Les trois études menées ont pour but de donner les clefs permettant l'analyse de l'intégrité du signal lors d'un empilement 3D présentant au moins deux niveaux.

- La première concerne la stratégie de routage où les performances de deux chaînes d'interconnexions sont comparées. Elles sont chacune constituées d'un Cu-Pillar, d'une interconnexion de RDL, d'un TSV et d'une interconnexion de BEOL. Dans la première chaîne, le routage est principalement effectué par une interconnexion de BEOL, alors que la seconde chaîne utilise principalement une interconnexion de RDL.
- La seconde étude concerne la stratégie d'orientation à adopter lors de l'empilement de trois niveaux. Cette étude donne l'orientation de la puce centrale à privilégier, en fonction de l'application visée. Pour cela deux cas sont étudiés. Dans le premier cas, la puce centrale est orientée en Face to Face avec la puce supérieure, alors que dans le second cas, la puce centrale est orientée en Face to Back avec la puce supérieure.
- La dernière étude porte sur la stratégie technologique, elle met en évidence cinq chaînes d'interconnexions composées d'éléments similaires, mais ayant des densités d'intégration différentes. Cette étude permet de déterminer les éléments critiques de chaque chaîne.

A partir de la matrice ABCD d'une chaîne d'interconnexion et, connaissant les drivers d'entrée et de sortie, il est donc possible de déterminer les performances de chaque scénarios en terme de retard à 50%.

Nous allons dans un premier temps expliciter les procédures mises en place permettant de mener nos analyses, en particulier celle utilisée pour quantifier les performances temporelles des interconnexions en partant des modèles équivalents définis dans le domaine fréquentiel.

## II Conversion du domaine fréquentiel au domaine temporel

L'évaluation des performances d'une interconnexion ou d'un réseau d'interconnexions intégrés s'établie généralement dans le domaine temporel dès lors que les signaux transmis sont de types numériques, qu'il s'agisse de prévision (à partir de simulation) ou de spécification (à partir de mesures).

La traduction des données dans le domaine temporel consiste à construire les réponses temporelles des réseaux d'interconnexions étudiés, sous forme de réponses impulsionnelles, à partir de données fréquentielles discrètes et finies.

Les principales étapes intervenant dans le passage du domaine fréquentiel au domaine temporel sont synthétisées par l'organigramme suivant :

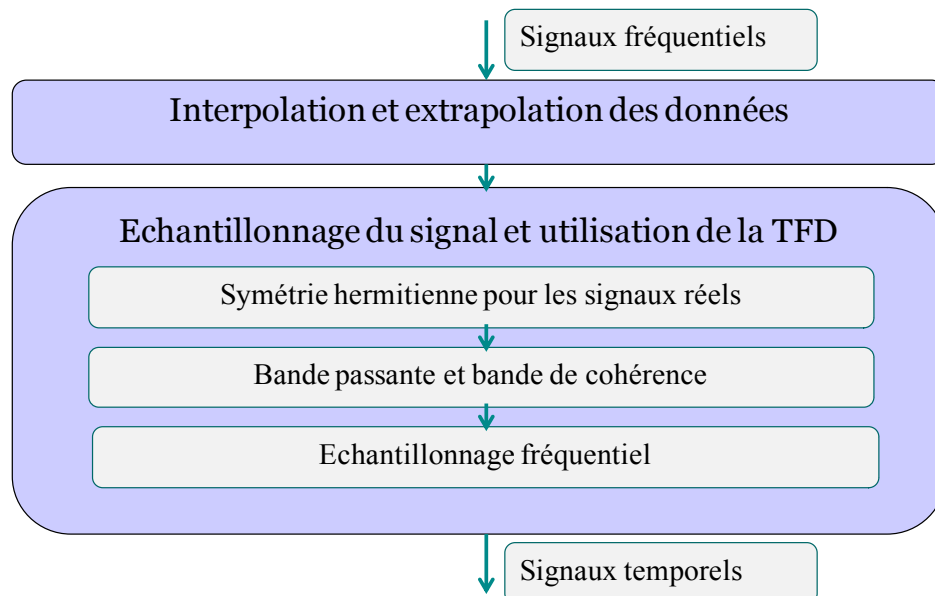


Figure 4-1 : Etapes du passage du domaine fréquentiel au domaine temporel.

L'interpolation : Consiste à ajouter des valeurs supposées entre deux données fréquentielles connues. En augmentant la résolution fréquentielle on garantit que la réponse impulsionnelle sera suffisamment longue.

L'extrapolation : Opération délicate, permettant d'ajouter des valeurs supposées au-delà de l'intervalle de données fréquentielles connues. Elle permet d'améliorer la résolution temporelle des signaux numériques calculés

La symétrie hermitienne pour signaux réels : Cette étape, nécessaire à l'application d'une TFD (Transformée de Fourier Discrète), consiste à calculer les valeurs de la réponse pour les fréquences négatives.

La bande passante et la bande de cohérence : Servent à appliquer le principe de Shannon-Nyquist duquel dépend la totalité et l'intégrité de l'information prélevée, lors d'un échantillonnage temporel.

L'échantillonnage fréquentiel : Si la bande de cohérence du signal mesuré est inférieure à la précision fréquentiel de la mesure, alors le signal est échantillonné sans dégradation de l'information. On mesure dans le domaine fréquentiel N valeurs discrètes qui décrivent la réponse en fréquence d'un système. Les caractéristiques des mesures effectuées sont les suivantes :

- $N = 1600$  points
- $F_{\text{prec}} = 25$  MHz
- $F_{\text{min}} = 40$  MHz
- $F_{\text{max}} = 40\,000$  MHz

La précision temporelle  $T_{\text{prec}}$  est donnée par :  $T_{\text{prec}} = 1/2.F_{\text{max}}$

La durée potentiel d'observation  $T_{\text{max}}$  est donnée par :  $T_{\text{max}} = 1/F_{\text{prec}}$

Remarque : Les principales étapes permettant l'obtention de réponses temporelles viennent d'être synthétisées. Cependant en pratique il faut également tenir compte d'autres problématiques comme la terminologie sur les signaux et les systèmes, les problèmes d'apodisation ainsi que de causalité.

### III Stratégie de routage

Lors de l'empilement de deux puces, le signal peut être routé sur un même plan soit grâce à une interconnexion de RDL, soit par une interconnexion de BEOL. Cependant, selon la taille des puces, les longueurs d'interconnexion peuvent varier de quelques microns à plusieurs millimètres. Il est

donc indispensable de quantifier en terme de retard à 50% dans la transmission d'un signal, les performances obtenues par un routage effectué principalement par une interconnexion de RDL et celles obtenues par un routage effectué principalement par une interconnexion de BEOL.

### III.1 Description des scénarios de routage

Dans une configuration Face to Back, la communication entre deux puces est étudiée par l'intermédiaire de deux scénarios, l'objectif étant d'identifier la meilleure stratégie de routage en tirant le meilleur parti des interconnexions disponibles. Les dimensions et topologie des briques élémentaires d'interconnexions, présentes ici, ont été choisies en regard des technologies actuellement matures et en accord avec les celles pressenties dans de futures applications.

Dans le premier cas (Figure 4-2-a), la communication entre les deux puces s'effectue, en terme de longueur, principalement par une interconnexion du BEOL. Depuis la puce supérieure on débute par l'intermédiaire d'un Cu-Pillar-d20, d'une interconnexion de RDL-w10 de longueur 50  $\mu\text{m}$ , d'un TSV-d10 et d'une interconnexion de BEOL (de longueur variable dans notre étude). Dans ce cas, le TSV est relativement éloigné du transistor, car c'est l'interconnexion de BEOL qui relie ces deux éléments.

Dans le second cas (Figure 4-2-b), cette même communication entre les deux puces s'effectue cette fois principalement par une ligne de RDL. Depuis la puce supérieure on débute par l'intermédiaire d'un Cu-Pillar-d20, d'une interconnexion de RDL-w10 (de longueur variable dans notre étude), d'un TSV-d10 et d'une interconnexion de BEOL de longueur 50  $\mu\text{m}$ . Dans ce cas, le TSV est relativement proche du transistor (à 50 $\mu\text{m}$ ).

Remarque : La position du TSV par rapport à celle d'un transistor a son importance. En effet, de nombreuses études sont actuellement en cours dans le but de quantifier l'impact électrique d'un TSV sur un transistor [Rousseau, 2009].

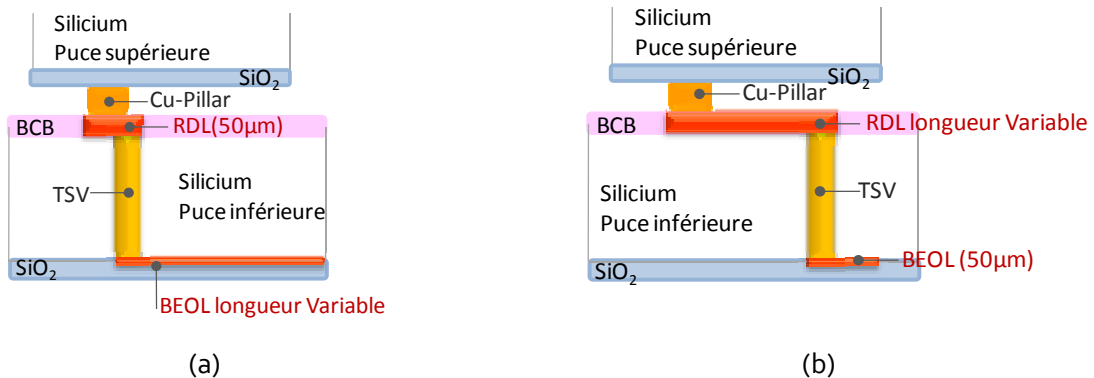


Figure 4-2 : Schéma des deux scénarios de routage (a) Routage principalement effectué par l'interconnexion de BEOL (b) Routage principalement effectué par l'interconnexion de RDL.

La comparaison des deux scénarios de routage se fait à longueur totale d'interconnexion constante où la longueur d'interconnexion dans le BEOL du premier cas est identique à la longueur de RDL dans le second cas.

### III.2 Résultats

Le retard induit par notre chaîne interconnexion par chaque scénario, en fonction de la longueur totale d'interconnexion, est représenté sur les Figure 4-3-a. Le retard engendré par le scénario utilisant principalement une interconnexion de BEOL est noté BEOL et le retard engendré par le scénario utilisant principalement une interconnexion de RDL est noté RDL. Comme le laissaient présager les études détaillées dans le chapitre 3, le retard à 50% est réduit lorsque l'on utilise la couche de redistribution plutôt que l'interconnexion de BEOL pour router le signal, et ce d'autant plus que la longueur totale de l'interconnexion est élevée.

Le deuxième graphique (Figure 4-3-b) montre la diminution relative de ce retard en fonction de la longueur totale d'interconnexion lorsque l'on privilégie le routage par une interconnexion RDL plutôt que par une interconnexion BEOL. Un gain de 60 % sur le temps de retard à 50 % pour une longueur totale d'interconnexion de 2 mm peut être obtenu avec routage principalement réalisé par une RDL.

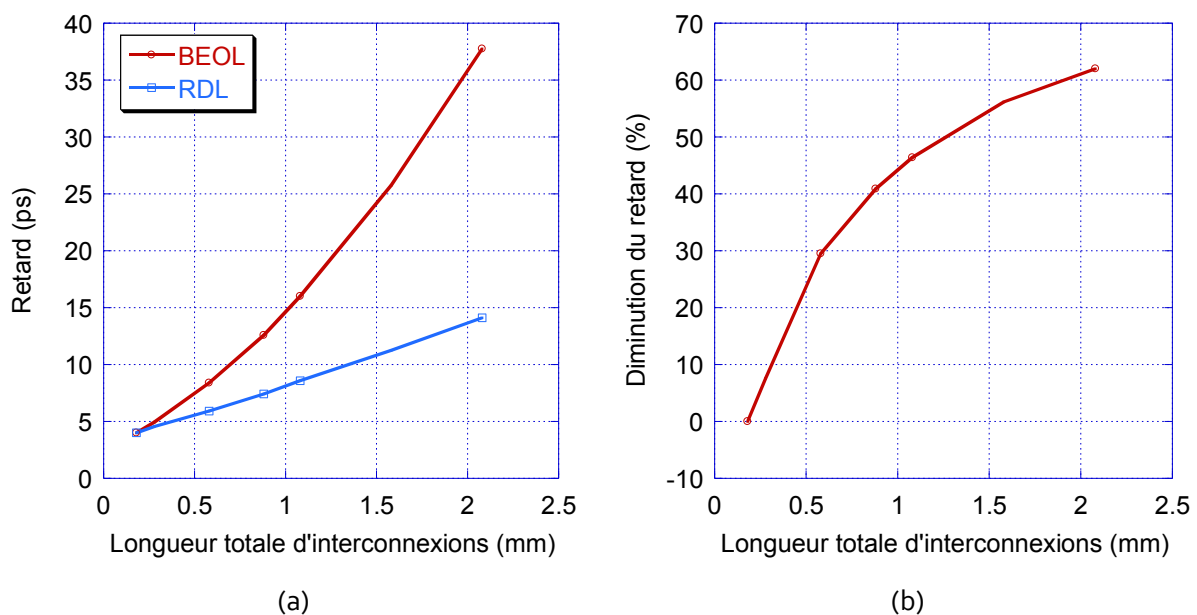


Figure 4-3 : (a) Retard en fonction de la longueur d'interconnexions (b) Diminution relative du retard lorsque l'on utilise un routage RDL plutôt qu'un routage BEOL.

Un tel écart sur le retard se comprend parfaitement en regardant le comportement fréquentiel de ces deux interconnexions. Les paramètres inductifs ( $0.7 \text{ pH}/\mu\text{m}$  pour la RDL et  $0.5 \text{ pH}/\mu\text{m}$  pour l'interconnexion de BEOL) ainsi que les paramètres capacitifs ( $0.35 \text{ fF}/\mu\text{m}$  pour la RDL et  $0.1 \text{ fF}/\mu\text{m}$  pour l'interconnexion de BEOL) restent relativement dans les mêmes ordres de grandeur. Cependant

le paramètre résistif de l'interconnexion de RDL ( $1.5 \text{ m}\Omega/\mu\text{m}$  en basse fréquence) est 40 fois moins élevé que celui de l'interconnexion de BEOL ( $61 \text{ m}\Omega/\mu\text{m}$  en basse fréquence). Ceci explique pourquoi une interconnexion de BEOL engendre plus de retard qu'une interconnexion de RDL.

Critère d'optimisation sur le routage : Plus les longueurs d'interconnexion nécessaires au routage sont élevées, plus il est rentable d'utiliser des interconnexions de RDL plutôt que des interconnexions de BEOL, pour minimiser les retards à 50 % dans la transmission d'un signal rapide entre deux puces empilées en Face to Back. Un gain de 60 % sur ce retard peut être atteint pour une longueur totale d'interconnexion de 2 mm.

## IV Stratégie d'orientation

Lors d'un empilement sur trois niveaux, par exemple dans le cas de trois puces empilées, le choix de l'orientation de la puce centrale devient primordial. Dans de nombreuses applications la puce centrale est généralement amenée à communiquer avec les deux autres puces à des vitesses très différentes, dont l'une peut être très élevée (plus d'une dizaine de GBy/s).

Dans cette partie, les performances en termes de retard à 50% lors des différentes communications entre deux puces empilées, elles mêmes empilées sur un substrat BGA, seront étudiées à travers deux scénarios d'orientation pour la puce située au centre.

### IV.1 Description des scénarios d'orientation des puces

La dénomination des deux cas étudiés est faite en fonction de la configuration d'empilement des deux puces supérieures. Le premier cas est nommé Face to Face (Figure 4-4-a) et le second cas est nommé Face to Back (Figure 4-4-b).

Dans les deux cas, la communication entre chaque niveau est étudiée pour comparer la propagation du signal puce à puce et puce à substrat.

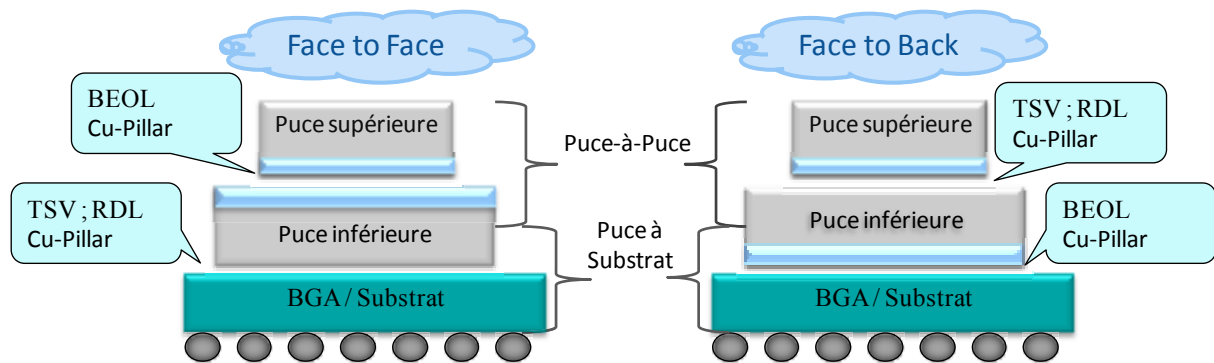


Figure 4-4 : Vue schématique des deux orientations possibles au cours d'un empilement à trois niveaux (a) Configuration Face to Face (b) Configuration Face to Back

Etant donné que l'orientation de la puce située au centre (nommée puce inférieure dans le schéma) diffère, la communication entre les deux niveaux n'est pas établie avec les mêmes composants d'interconnexion. Pour l'orientation Face to Face, la communication puce à puce se s'effectue par une interconnexion de BEOL et d'un Cu-pillar-d20 alors que la communication puce à substrat est effectuée par un TSV-d10, une interconnexion de RDL-W10 et Cu-pillar-D55. Pour l'orientation Face to Back, la communication puce à puce se compose d'un Cu-pillar-d20, d'un TSV-d10 et d'une RDL-w10 alors que la communication puce à substrat est effectuée par une interconnexion de BEOL et un Cu-pillar-D55.

## IV.2 Résultats

La Figure 4-5 montre le retard induit par chaque chaîne d'interconnexion en fonction de la longueur totale de la chaîne.

Le fait que les courbes de résultats soient superposées pour le retard à 50 % sur la communication puce à puce (configuration Face to Face) et la communication puce à substrat (configuration Face to Back) ainsi que la superposition des courbes de retard pour la communication puce à puce (configuration Face to Back) et la communication puce à substrat (configuration Face to Face) montre que l'influence sur le retard des Cu-Pillar est négligeable.

En effet, les éléments constituant la communication puce à puce (configuration Face to Face) sont presque identiques à ceux constituant la communication puce à substrat (configuration Face to Back), la seule différence étant la nature des Cu-Pillar. Il en est de même en ce qui concerne la communication puce à puce (configuration Face to Back) et la communication puce à substrat (configuration Face to Face).

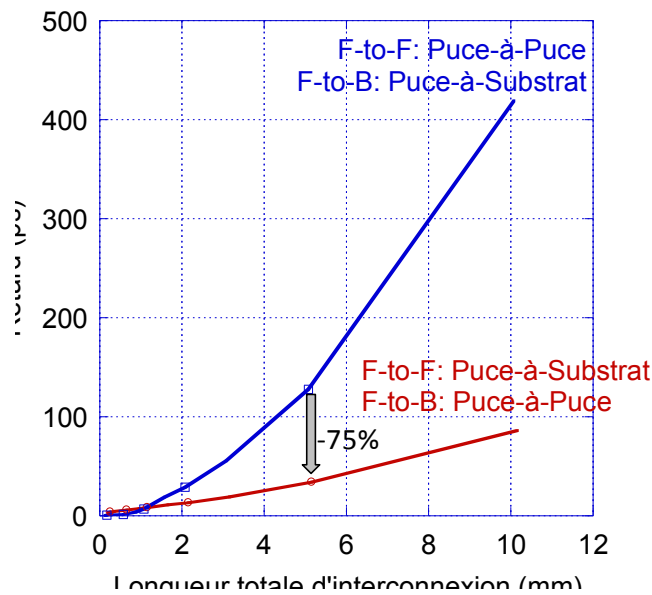


Figure 4-5 : Retard en fonction de la longueur d'interconnexion

Ces résultats démontrent clairement que le choix d'une configuration Face to Face ou Face to Back doit être examiné avec précaution, et cela en fonction de la communication la plus exigeante à en terme de rapidité de communication entre les liaisons puce à puce ou puce à substrat. Pour une liaison puce à puce la configuration Face to Back réduit les temps de propagation alors que pour une liaison puce à substrat c'est la configuration Face to Face qui réduit de ce temps de propagation.

D'après la Figure 4-5, ces choix d'optimisation permettent de réduire le retard de 75 % pour des interconnexions de longueur totale de 5 mm.

Critère d'optimisation sur l'orientation: Pour une optimisation de la rapidité de communication puce à puce, la meilleure configuration est le Face to Back. Par contre, si l'optimisation doit se faire sur la communication puce à substrat alors il faudra privilégier une configuration Face to Face. Une diminution de 75 % sur le retard peut être atteinte pour des liaisons longues (5 mm).

## V Stratégie technologique.

Dans cette dernière partie, les temps de retard de plusieurs chaînes d'interconnexions 3D en configuration Face to Back composées des mêmes éléments mais de densité d'intégration différentes sont étudiées. Après avoir décrit les briques élémentaires d'interconnexion composant chaque chaîne, l'impact respectif de chacune de ces briques élémentaires sur le retard sera détaillé. La densification progressive des interconnexions 3D nécessite de nombreux développements technologiques. L'objectif est ici d'évaluer leur apport en termes de performance.



## V.1 Description des scénarios technologiques

Lors de la description des briques élémentaires dans le chapitre 3, cinq types de TSV ont été détaillés. Chacun des TSV diffère des autres par sa densité d'intégration, conditionnant leur topologie, type de remplissage ou de processus de fabrication. Pour former une chaîne d'interconnexion, comme représenté en Figure 4-6, chaque TSV est associé à un Cu-Pillar (liaison puce à puce), à une interconnexion de RDL et à un second Cu-Pillar (liaison puce à substrat), (cf Chapitre 1, partie III-1-4).

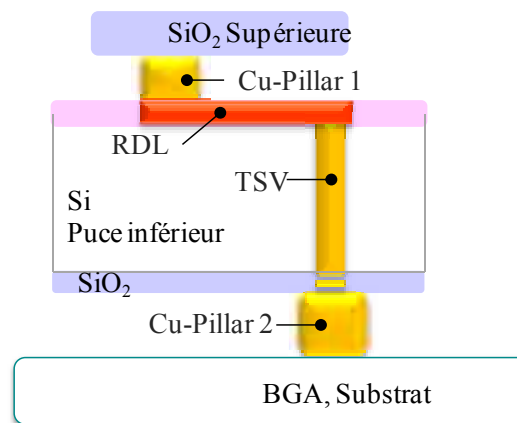


Figure 4-6 : Vue schématique de la chaîne complète.

Afin d'être cohérent en terme de densité d'intégration, chaque TSV est associé à des éléments de densité d'intégration similaire.

Les cinq chaînes d'interconnexion 3D étudiées sont présentées dans le tableau ci-dessous.

Tableau 5-1 : Chaines d'interconnexions 3D étudiées, de densité faible à densité haute				
	<i>Cu-Pillar 1</i>	<i>RDL</i>	<i>TSV</i>	<i>Cu-Pillar 2</i>
<i>Chaîne 1</i>	d20	w20	d60	d75
<i>Chaîne 2</i>	d20	w10	d10	d55
<i>Chaîne 3</i>	d20	w10	d6	d55
<i>Chaîne 4</i>	d20	w10	d5	d55
<i>Chaîne 5</i>	d3	damas	d3	d35

## V.2 Résultats

### V.2.1 Performance en termes de retard.

Le retard induit par chaque chaîne d'interconnexion, relative à une densité d'intégration, est représenté en Figure 4-7 en fonction de la longueur totale d'interconnexion. Les Cu-Pillar et les TSV ayant des dimensions fixes, seule la longueur de l'interconnexion de RDL varie.

Les chaînes 2, 3 et 4 apportent des réponses similaires sur le temps de retard à 50 %. Toutes les trois font partie de la même densité d'intégration et par conséquent les Cu-Pillar et les interconnexions de RDL utilisés sont identiques et seules les dimensions des TSV varient. Cela prouve que l'écart observé sur la Figure 4-7-a n'est dû qu'à la différence de densité d'intégration des TSV. Le TSV-d5 est moins résistif et moins capacitif que les TSV-d10 et d6. Ceci explique la meilleure performance de la chaîne 4. Pour une longueur d'interconnexion égale à 2 mm, le retard est réduit de 14% en choisissant TSV-d5 plutôt qu'un TSV-d10.

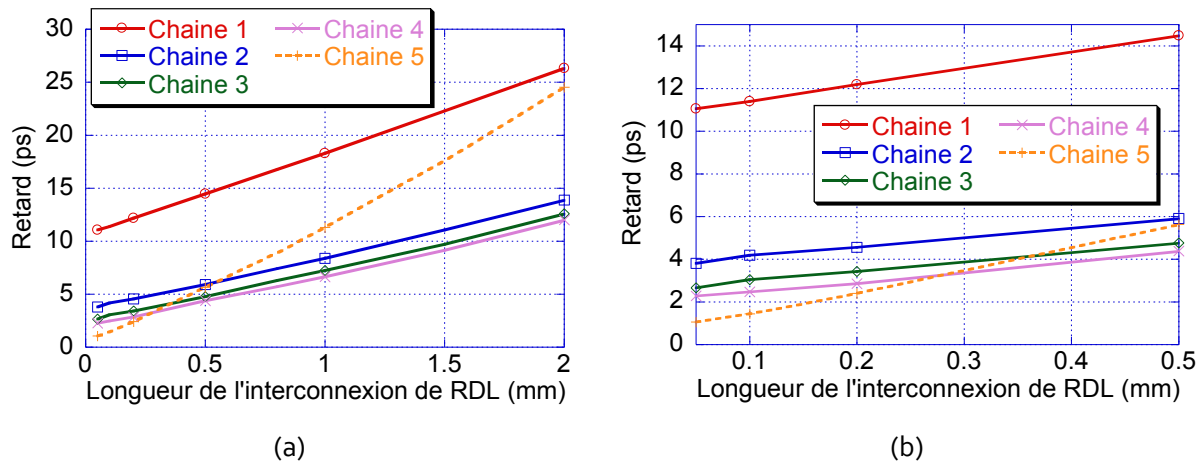


Figure 4-7 : Retard de chaque chaîne en fonction de la longueur d'interconnexion.  
(a) Pour des longueurs d'interconnexion comprises entre 50 µm et 2 mm. (b) Pour des longueurs d'interconnexion comprises entre 50 et 500 µm.

Les chaînes 1 et 5 ne peuvent être vraiment comparées aux autres car elles ont des densités d'intégration différentes et ne visent donc pas les mêmes applications. Néanmoins, la chaîne 1 entraîne des retards à 50 % élevés du fait de la nature capacitive du TSV-d60. La chaîne 5 reste intéressante, spécialement pour de courte longueur d'interconnexion (Figure 4-7-b). En effet lorsque la longueur d'interconnexion augmente, la nature hautement résistive de la RDL-damas augmente de manière significative le retard.

## V.2.2 Influence relative de chaque brique élémentaire d'interconnexion dans une chaîne complète d'interconnexion.

A partir de la Figure 4-8 l'élément de la chaîne qui influence le plus le retard peut être déterminé en fonction de la longueur de l'interconnexion de RDL (la Figure 4-8-a fait référence à la chaîne 1 et la Figure 4-8-b à la chaîne 5). L'objectif est d'identifier pour chaque génération technologique

d'intégration 3D la brique élémentaire sur laquelle les concepteurs et technologues doivent focaliser leurs efforts pour améliorer les performances.

En ce qui concerne la chaîne 1 (faible densité d'intégration), on remarque que lorsque la longueur de l'interconnexion de RDL est inférieure à 1,4 mm le retard est principalement dû au TSV. Au delà de cette valeur, c'est l'interconnexion de RDL qui devient l'élément provoquant le plus de retard. La longueur pour laquelle le retard est provoqué équitablement par le TSV et l'interconnexion de RDL est nommé « longueur critique ». En dessous de cette valeur le TSV est l'élément critique alors qu'au dessus c'est l'interconnexion de RDL.

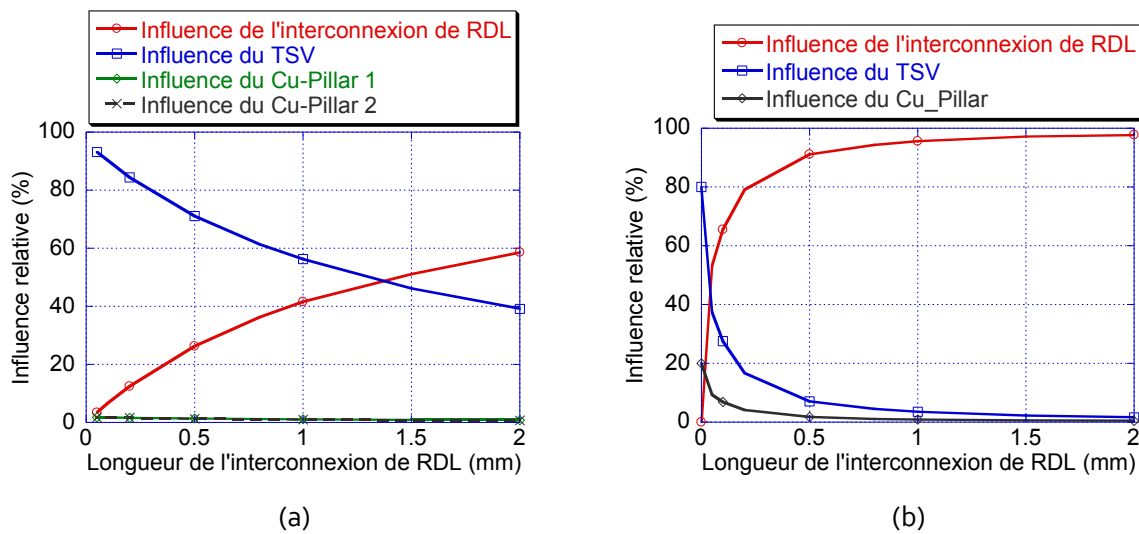


Figure 4-8 : Influence relative de chaque brique élémentaire sur le retard en fonction de la longueur d'interconnexion. (a) Chaîne 1. (b) Chaîne 5.

Pour la chaîne 5 (haute densité d'intégration), la longueur critique est notablement plus faible  $\approx 40 \mu\text{m}$ . En effet, le TSV haute densité induisant un retard faible, l'interconnexion de RDL devient rapidement l'élément critique. Dans les deux cas, la part de retard que provoquent les Cu-Pillars reste négligeable.

La longueur critique varie en fonction des différents types de chaîne d'interconnexion étudiés (Figure 4-9) et elle donne une indication quant à la nature de l'élément le plus critique lors de la transmission d'un signal.

Concernant les chaînes 2, 3 et 4, la longueur critique est obtenue lorsque la longueur d'interconnexion de RDL est comprise entre  $400 \mu\text{m}$  et  $800 \mu\text{m}$ . Avant ces valeurs, le retard est principalement dû au TSV, au-delà l'interconnexion de RDL-W10 devient l'élément le plus critique. A partir de cette figure, on constate que plus la densité d'intégration est importante, plus la longueur de l'interconnexion de RDL diminue pour atteindre la longueur critique.

Critère d'optimisation sur les densités d'intégration : Plus la densité d'intégration augmente, plus l'interconnexion de RDL devient rapidement l'élément critique, et ce d'autant plus pour des longueurs élevées. Cependant pour des interconnexions de RDL relativement courtes, inférieures à 1,4 mm, pour de faibles densités d'intégration ou inférieures à 400  $\mu\text{m}$ , pour les moyennes densités d'intégration, c'est le TSV qui est l'élément à optimiser.

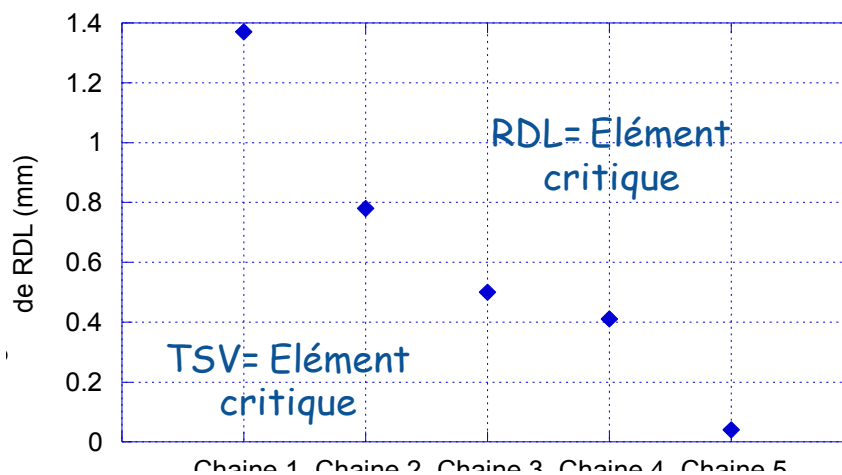


Figure 4-9 : Longueur de l'interconnexion de RDL pour laquelle le TSV et l'interconnexion de RDL ont le même impact sur le retard, en fonction des différentes chaînes d'interconnexion

## VI Conclusion

Toutes les interconnexions spécifiques à la technologie d'intégration 3D ont été étudiées et modélisées dans le chapitre 3. L'association de ces briques élémentaires en fonction de différentes stratégies d'intégration donnent quelques règles majeures à suivre, que ce soit en termes de routage, d'orientation ou encore d'option technologique.

Les interconnexions de RDL offrent de meilleures performances de routage par rapport à une interconnexion de BEOL pour des distances supérieures à 1,4 mm grâce à l'isolation low-k et à la forte épaisseur de métallisation. Plus les longueurs d'interconnexion nécessaires au routage sont élevées, plus il est rentable d'utiliser des interconnexions de RDL plutôt que des interconnexions de BEOL. Une diminution de 60 % sur ce retard peut être atteinte pour une longueur totale d'interconnexion de 2 mm.

L'orientation de la puce centrale, lors d'un empilement à trois niveaux doit résulter d'une analyse claire des spécifications du produit afin d'optimiser au mieux une communication puce à puce ou puce à substrat, en fonction d'une configuration Face to Face ou Face to Back. Pour une optimisation de la rapidité de communication puce à puce, la meilleure configuration est le Face to Back. Par contre, si l'optimisation doit se faire sur la communication puce à substrat alors il faudra

privilégier une configuration Face to Face. Une diminution de 75 % sur le retard peut être atteinte pour des liaisons longues (5 mm).

L'élément critique dans une chaîne d'interconnexion dépend principalement de la densité d'intégration. La loi de Moore est également valable en intégration 3D. Ces recommandations générales constituent une première étape pour aider les concepteurs à trouver les meilleurs scénarios d'intégration 3D pour leurs produits. Plus la densité d'intégration augmente, plus rapidement l'interconnexion de RDL devient l'élément critique, et ce d'autant plus pour des longueurs élevées. Cependant pour des interconnexions de RDL relativement courtes, inférieures à 1,4 mm, pour de faibles densités d'intégration ou inférieures à 400  $\mu\text{m}$ , pour les moyennes densités d'intégration, c'est le TSV qui est l'élément à optimiser.

Tous ces éléments nous permettent de mieux appréhender les performances possibles pour des applications réalistes, en fonction de leurs designs, de leurs empilements ou encore de leurs densités d'intégration.

Le chapitre suivant prendra l'exemple d'une application concrète, pressentie comme l'application principale de l'intégration 3D, l'assemblage d'une mémoire et d'un processeur : le Wide I/O.



# Chapitre 5

## Evaluation des performances et stratégies d'empilements 3D pour des circuits rapides de type mémoire- processeur (Wide I/O)

---

A l'heure actuelle, le nombre de nouvelles applications offertes par l'intégration 3D et les performances prometteuses qu'il serait possible d'atteindre justifient les débats ouverts quant à de profonds changements (et lourds investissements) pour mettre en place les nouveaux procédés de fabrication qui lui sont nécessaires. En effet empiler les puces, les unes sur les autres, pour éviter les mises en boîtiers intermédiaires implique une révolution dans les procédés de fabrication. L'amincissement des puces à quelques dizaines microns, le recours à des techniques de collage intermédiaire ou encore la conception de TSV, ne sont que quelques exemples d'innovations technologiques inhérente à l'intégration 3D.

La réalisation d'objets toujours plus petits, toujours plus rapides, toujours plus avides de mémoires et toujours moins gourmands en énergie est l'objectif à atteindre dans les applications de l'intégration 3D. L'application qui semble répondre à ces attentes est l'association d'une mémoire de type Wide I/O avec un processeur performant.

Ce chapitre, à l'aide des résultats obtenus précédemment, démontre les possibilités d'associer une mémoire Wide I/O avec un processeur suivant différents empilements, différentes plateformes technologiques et différentes densité d'intégration. Pour chaque configuration les spécifications en termes de fréquence maximale de fonctionnement et d'intégrité des signaux seront étudiées.

Dans un premier temps nous allons étudier le cas des empilements Face to Face et Face to Back, sans interposer, afin de déterminer les vitesses de fonctionnement envisageables de chaque type de communication, entre les puces, entre les puces et le substrat BGA. Puis nous les comparons aux performances obtenues avec un empilement comportant un interposer. Ces études sont réalisées avec des briques élémentaires d'interconnexions faisant propres à la technologie de moyenne densité d'intégration. Les TSV utilisés sont de diamètre 10  $\mu\text{m}$  et les piliers de cuivre, nécessaires à la communication Processeur – BGA, de diamètre 55  $\mu\text{m}$ .

Afin de couvrir une majorité de configuration, la dernière partie reprend, sous forme d'abaque, les résultats obtenus pour trois types empilements et présente les performances attendues sur chaque type de communication et pour quatre densités d'intégration différentes. Préalablement la démarche détaillée de notre analyse sera exposée pour un cas particulier, le Face to Back sans interposer.

Pour finir une stratégie d'intégration 3D répondant aux spécifications des Wide I/O sera déterminée pour chaque type d'empilement.



## Sommaire

<b>I</b>	<b>INTRODUCTION.....</b>	<b>156</b>
I.1	SPECIFICATIONS POUR CIRCUIT MEMOIRE-LOGIQUE. ....	158
I.2	LES ALTERNATIVES D'EMPILEMENT. ....	159
<b>II</b>	<b>COMMUNICATION MEMOIRE-LOGIQUE ET LOGIQUE-BGA (AVEC ET SANS INTERPOSER)....</b>	<b>160</b>
II.1	EMPILEMENTS SANS INTERPOSER .....	160
II.1.1	<i>Description des empilements .....</i>	<i>160</i>
II.1.2	<i>Calcul des fréquences maximales de communication. ....</i>	<i>160</i>
II.2	L'INTERPOSER : UNE SOLUTION DE PLATEFORME TECHNOLOGIQUE 3D PROMETTEUSE.....	163
II.2.1	<i>Description de l'empilement pour une solution avec Interposer .....</i>	<i>163</i>
II.2.2	<i>Résultats obtenus sur les performances électriques.....</i>	<i>164</i>
II.2.2.a	Fréquences maximales des signaux .....	165
II.2.2.b	Etude du retard en fonction des longueurs de RDL.....	166
II.2.2.c	Etude de l'intégrité des signaux à l'aide de diagramme de l'œil. ....	167
II.3	CONCLUSION DES PARTIES 2 ET 3. ....	169
<b>III</b>	<b>COMMUNICATIONS MEMOIRE-PROCESSEUR ET PROCESSEUR-BGA EN FONCTION DES DENSITES D'INTEGRATION (AVEC ET SANS INTERPOSER).....</b>	<b>170</b>
III.1	CAS DE L'EMPILEMENT FACE TO BACK .....	170
III.1.1	<i>Description de l'empilement.....</i>	<i>170</i>
III.1.2	<i>Résultats sur l'empilement Face to Back.....</i>	<i>171</i>
III.1.2.a	Fréquence des signaux .....	171
III.1.2.b	Intégrité des signaux - diagrammes de l'œil. ....	172
III.2	ROADMAP ET STRATEGIE D'EMPILEMENT POUR LES CIRCUITS ASSOCIANT UNE MEMOIRE A UN PROCESSEUR.....	174
III.2.1	<i>Fréquences maximales en fonction des générations technologiques. ....</i>	<i>174</i>
III.2.2	<i>Stratégies d'empilement selon le type de densité d'intégration.....</i>	<i>175</i>
<b>IV</b>	<b>CONCLUSION .....</b>	<b>176</b>

## I Introduction

Additionnées à des processeurs multimédia, les mémoires sont des composants clés de toute innovation technologique. Les performances graphiques des nouveaux outils de communication comme les Smartphones ou les tablettes provoquent de nombreuses attentes principalement au niveau de la bande passante et de la consommation. En effet l'autonomie d'un Smartphone illustre bien cette demande, à l'heure actuelle elle ne dépasse pas la journée, contrairement à celle d'un téléphone classique qui avoisine les 10 jours. Bien évidemment cela dépend du mode d'utilisation, mais globalement, la contrainte utilisateur est forte.

Les mémoires faible puissance, les plus utilisées à l'heure actuelle sont les LPDDR2. Elles offrent des fonctionnalités avancées de gestion de l'alimentation, une interface révolutionnaire partagée pour la mémoire non volatile (NVM), de la mémoire volatile (SDRAM) et une large gamme de densités et de vitesses. Cependant leur utilisation est limitée par des problèmes thermiques et par une trop grande consommation, pour un gain en bande passante assez relatif. Elles peuvent avoir une bande passante maximale de 8,5 GBy/s pour une consommation de 800 mW (Figure 5-1).

La volonté d'améliorer les performances des LPDDR2 a provoqué l'arrivée des mémoires LPDDR3. Actuellement en développement, la norme LPDDR3 offrira une extension à la LPDDR2, en termes de bande passante, atteignant 12,8 GBy/s pour une configuration dual channel. La LPDDR3 est conçue pour répondre aux besoins des nouveaux Smartphones et tablettes nécessitant plus de bande passante. Cette voie d'amélioration n'est tout de même pas viable à long terme. En effet même si la bande passante peut atteindre 12,8 GBy/s la consommation est très élevée, de l'ordre de 1W, limitant les possibilités d'évolution.

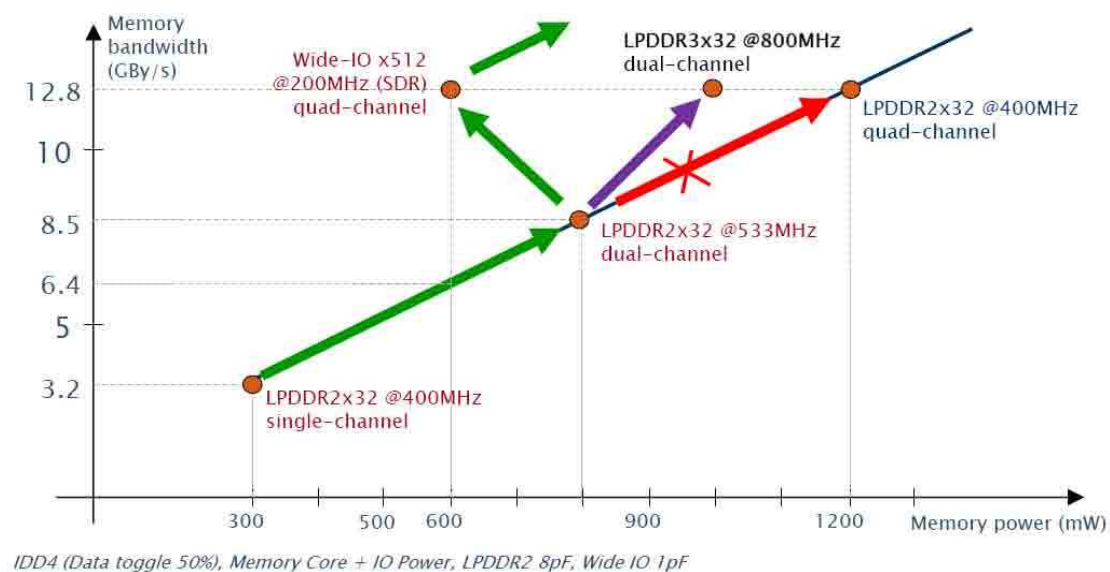


Figure 5-1 : Bande passante et consommation énergétique pour différentes applications mémoires.

Ce constat force les concepteurs à se tourner vers un nouveau type de mémoire, nommé Wide I/O.

Toujours en développement (Figure 5-2), le Wide I/O est une technologie révolutionnaire qui va répondre aux demandes de l'industrie pour des niveaux accrus d'intégration, avec un facteur de performance en bande passante et en puissance amélioré. Le Wide I/O DRAM est un niveau du circuit 3D relié avec le processeur au moyen de piliers de cuivre et de TSV. Il est particulièrement adapté pour des applications nécessitant une forte bande passante, jusqu'à 12,8 GBy/s et une consommation réduite à 600 mW, tels que les jeux 3D, vidéo HD ou les applications simultanées [JEDEC, 2011].

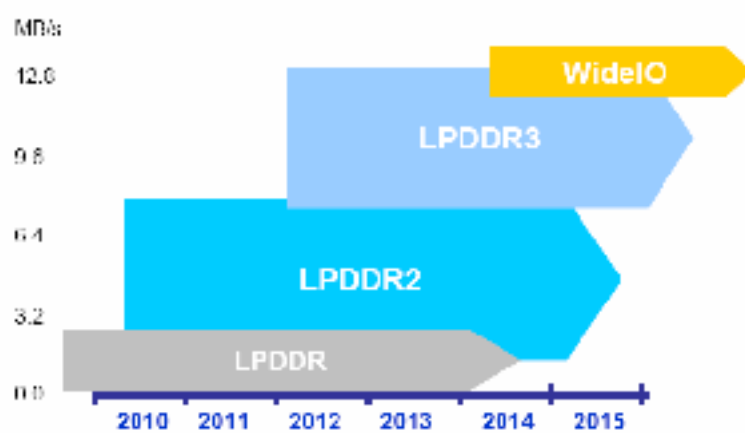


Figure 5-2 : Année de commercialisation des différentes mémoires mobiles.

En effet, là où la mémoire DDR envoie des informations à très haute fréquence (provoquant des problèmes d'échauffement et de consommation) sur un nombre restreint de canaux de communication, le Wide I/O multiplie le nombre de canaux de communication, permettant un envoi d'information supérieur à celui d'une mémoire DDR, à des fréquences plus faibles. Pour une bande passante de 12,8 GBy/s, les fréquences de communication des mémoires LPDDR2 et LPDDR3 sont respectivement de 400 et de 800 MHz contre seulement 200 MHz pour une mémoire Wide I/O.

L'innovation majeure de cette mémoire réside principalement dans son type d'intégration. C'est une technologie toute intégrée où la communication entre les puces empilées se fait par l'intermédiaire de TSV. Ce type d'intégration évite au Wide I/O d'utiliser une architecture dite « package on package » (POP), utilisé par exemple pour des mémoires de type LPDDR2 et LPDDR3, qui relie des mémoires à des processeurs, préalablement mis en boîtier, par l'intermédiaire d'une interface BGA.

La Figure 5-3 donne les pourcentages d'amélioration, en terme d'encombrement, de consommation d'énergie et de bande passante, obtenue par le passage d'une architecture dite « package on package » (POP) à une architecture avec TSV.

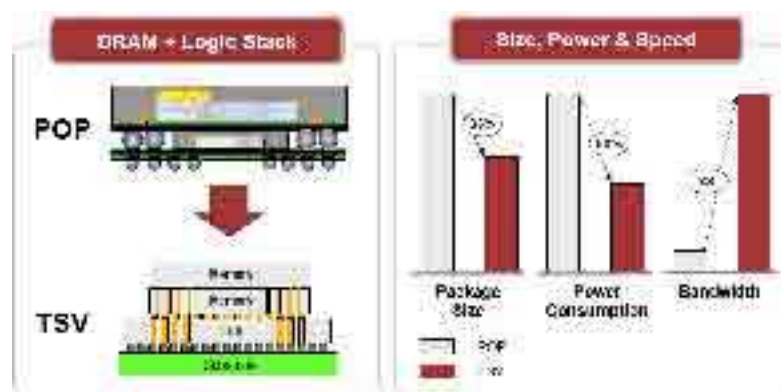


Figure 5-3 : Critères d'amélioration en passant d'une architecture POP à une architecture TSV.

On remarque qu'une intégration 3D avec TSV, en comparaison avec une architecture POP, diminue l'encombrement de 35%, la consommation d'énergie de 50% et permet de multiplier par huit la bande passante.

### I.1 Spécifications pour circuit Mémoire-Logique.

Prenons l'exemple d'une application Wide I/O, dans laquelle il est nécessaire d'empiler la puce mémoire sur le processeur multimédia et le processeur multimédia sur un BGA. Les spécifications en terme de fréquence d'entrée / sortie, sont drastiques (Figure 5-4). En effet, elles définissent à quelle vitesse maximale les deux puces communiquent.

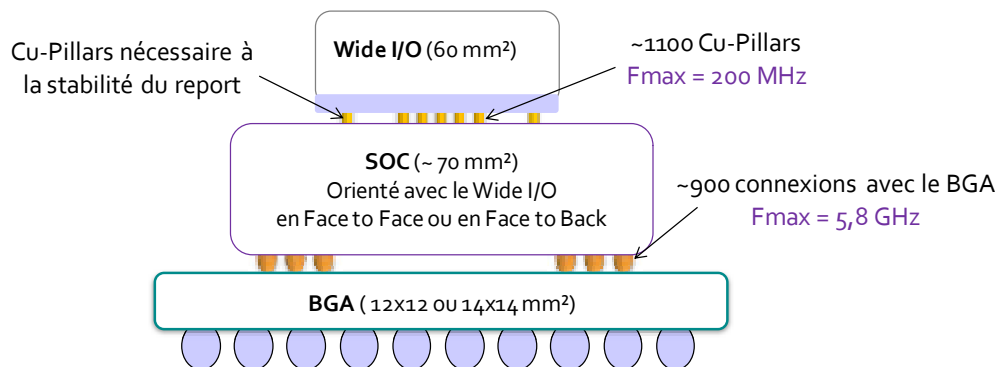


Figure 5-4 : empilement dans une application Wide I/O et fréquences de travail pour chaque communication entre puces et avec le BGA.

Pour cette application, la mémoire Wide I/O communique avec le processeur à une fréquence maximale de 200 MHz. Le processeur (SOC) communique, quant à lui avec le BGA à une fréquence maximale de 5,8 GHz. En ce qui concerne les autres types de mémoire, on a vu que les LPDDR2 et les LPDDR3 communiquent respectivement avec le processeur à des vitesses de 400 et 800 MHz.

## I.2 Les alternatives d'empilement.

En fonction des empilements, des plateformes technologiques ou encore des densités d'intégration les vitesses de communication entre une mémoire et un processeur et entre un processeur et un BGA doivent être déterminées. Les performances en termes de retards ainsi que la qualité des signaux, au travers de diagramme de l'œil, sont également des critères importants lors d'une étude sur la transmission de signaux. L'objectif de ce chapitre est donc de déterminer quelles sont les solutions technologiques les plus à même de répondre aux défis de l'intégration des mémoires Wide I/O.

Ce chapitre permettra de déterminer si les trois options illustrées Figure 5-5, remplissent les spécifications demandées en terme de bande passante et d'intégrité du signal. Dans un premier temps nous allons étudier le cas des empilements Face to Face et Face to Back, sans interposer, afin de déterminer les vitesses de chaque communication. Dans un deuxième temps nous les comparerons aux performances obtenues avec un empilement comportant un interposer.

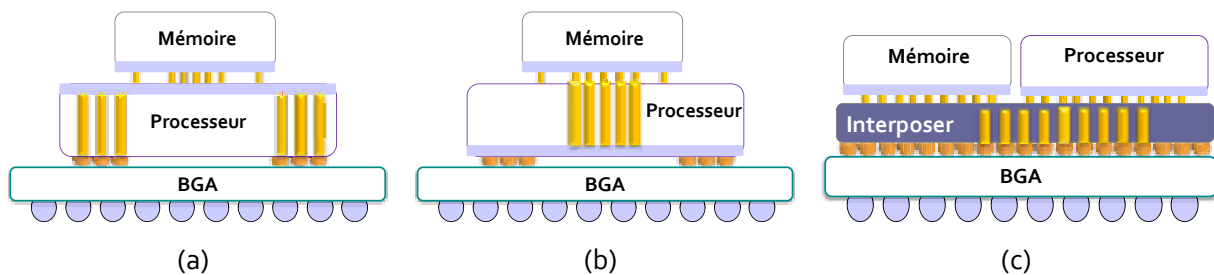


Figure 5-5 : Schémas illustratifs. (a) Orientation Face to Face. (b) Orientation Face to Back. (c) Utilisation d'un interposer.

Afin de généraliser nos études, la dernière partie reprend, sous forme d'abaque, les résultats obtenus pour trois types d'empilements et présente les performances attendues sur chaque type de communication et pour quatre densités d'intégration différentes.

La première densité d'intégration supposée est la moins intégrée. Elle utilise des TSV de diamètre 60  $\mu\text{m}$  et des piliers de cuivre de diamètre 75  $\mu\text{m}$ . La seconde et la troisième appartiennent à la moyenne densité d'intégration, avec respectivement des TSV de diamètre 10 et 6  $\mu\text{m}$  et des piliers de cuivre de diamètre 55  $\mu\text{m}$ . La dernière densité d'intégration étudiée est la plus agressive, elle utilise des TSV de diamètre 3  $\mu\text{m}$  et des piliers de cuivre de diamètre 35  $\mu\text{m}$ . Préalablement la démarche détaillée pour mener notre analyse sera exposée sur le cas particulier de l'empilement Face to Back sans interposer.

Pour finir une stratégie d'intégration 3D répondant aux spécifications des Wide I/O sera déterminée pour chaque type d'empilement.

## II Communication Mémoire-Logique et Logique-BGA (avec et sans interposer)

### II.1 Empilements sans interposer

Dans la partie suivante deux orientations pour l'empilement de la mémoire Wide I/O sur le processeur vont être étudiées. L'orientation entre la mémoire et le processeur est soit de type Face to Face soit de type Face to Back.

#### II.1.1 Description des empilements

Dans chaque empilement il existe deux chaînes d'interconnexions. Pour un empilement Face to Face (Figure 5-6-a), la mémoire est reliée au processeur par un Cu-Pillar et le processeur est relié au BGA par une interconnexion de BEOL ( $L=20\mu\text{m}$ ), un TSV et un Cu-Pillar. Dans le cas du Face to Back (Figure 5-6-b), la communication entre la mémoire et le processeur est effectuée par un Cu-Pillar, un TSV et une interconnexion de BEOL ( $L=20\mu\text{m}$ ) alors que la communication entre le processeur et le BGA se fait par l'intermédiaire d'un Cu-Pillar.

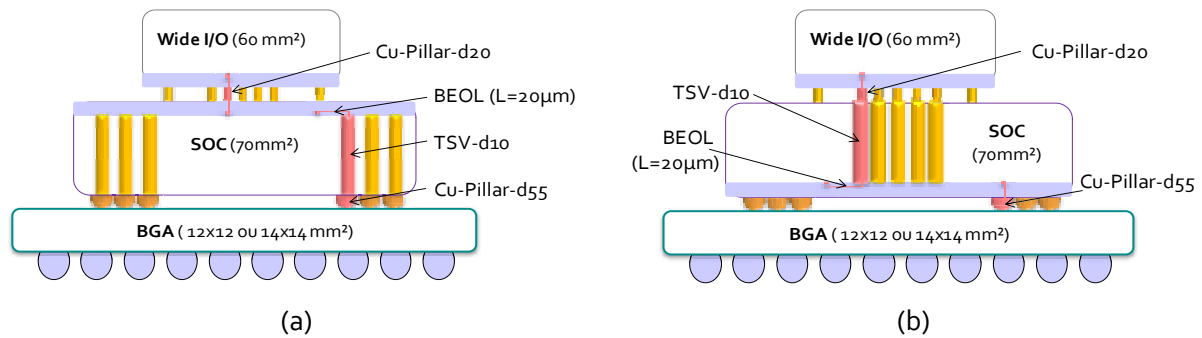


Figure 5-6 : Orientations entre la mémoire et le processeur. (a) Face to Face (b) Face to Back et description des chaînes d'interconnexions.

Dans cette partie, les briques élémentaires utilisées sont de moyenne densité d'intégration. Les TSV sont de diamètre 10  $\mu\text{m}$ , les Cu-pillar utilisés entre le processeur et la mémoire sont de diamètre 20  $\mu\text{m}$  et les Cu-pillar utilisés entre le processeur et le BGA sont de diamètre 55  $\mu\text{m}$ . Toutes les caractéristiques de ces éléments (géométrique, électrique...) ont été exposées dans les chapitres précédents.

#### II.1.2 Calcul des fréquences maximales de communication.

La fréquence maximale ( $F_{\text{Max}}$ ) de communication est intimement liée au temps de montée ( $T_m$ ) des signaux numériques en sortie de ligne

Il est généralement admis qu'un signal numérique, décrit par une succession périodique ( $T_c$ ) de niveau haut et bas est correctement transmis lorsque le temps de transition  $T_m$  (montée et descente) est au moins 4 fois plus faible que la période  $T_c$  du signal. De cette façon la fréquence maximale  $F_{\text{max}}$  du signal numérique est conditionnée par le temps de montée  $T_m$  selon la relation :

$$F_{Max} = 1 / (4 \cdot T_m)$$

Équation 5-1

Pour connaître cette fréquence maximale de communication sur une interconnexion il donc nécessaire de déterminer le temps de montée du signal transmis en sortie de l'interconnexion. Bien entendu le calcul de ce temps de montée en sortie de l'interconnexion n'est réalisable que si on connaît le temps de montée en entrée. L'obtention de ce dernier est possible par l'intermédiaire du logiciel MASTAR [Bœuf, 2007]. C'est un outil de calcul spécialement conçu pour l'évaluation des caractéristiques électriques des transistors CMOS, réalisés avec différentes technologies (*planar bulk*, Double Grille, Silicium sur isolant), constituant les drivers placés en entrée et sortie des réseaux d'interconnexions. Le calcul est basé sur des équations dépendant des principaux paramètres technologiques du transistor (longueur de grille, dopage de canal, épaisseur d'oxyde, etc.).

Outre son application principale, MASTAR est capable de calculer le temps de montée d'un échelon à l'entrée d'une chaîne d'interconnexion en fonction de la résistance et de la capacité équivalente de cette chaîne.

Une chaîne d'interconnexions est construite par l'association de chaque brique élémentaire d'interconnexion. En associant les paramètres résistifs et capacitifs de chaque brique, on obtient une résistance et une capacité globale (notée Rvar et Cvar) et variables en fonction de la fréquence. Dans le cas présent on souhaite obtenir des valeurs de R et de C équivalentes et constantes en fréquence, c'est pourquoi la fonction de transfert de la chaîne est dans un premier temps tracée avec les valeurs Rvar et Cvar, puis dans un second temps, un couple de constantes R et C donnant la même fonction de transfert est déterminé (Figure 5-7).

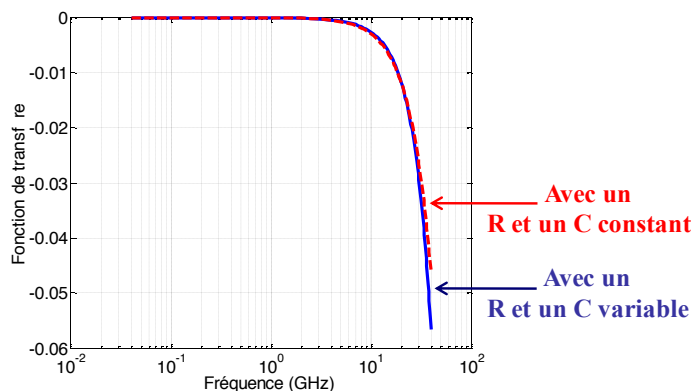


Figure 5-7 : Fonctions de transfert avec un couple RC variable et un couple RC constant.

Une fois que le couple RC équivalent d'une chaîne d'interconnexion est connu on obtient, par le logiciel MASTAR, le temps de montée des signaux à l'entrée de la chaîne d'interconnexion.

Nous déterminons ensuite le temps de montée en sortie de la chaîne d'interconnexion et nous en déduisons la fréquence maximale utilisable pour cette liaison.

Les résultats montrent que les communications Mémoire - Processeur effectuées avec les briques élémentaires citées précédemment, peuvent supporter des fréquences de communication très fortes, 11,9 GHz pour le Face to Face et 7,5 GHz pour le Face to Back. En ce qui concerne la communication Processeur - BGA, le phénomène est inversé, les orientations Face to Face et Face to Back, permettent respectivement que des fréquences maximales de 7,5 et 11,9 GHz (Tableau 1-1).

Tableau 1-1 : Fréquences maximales de communication : cas d'une moyenne densité d'intégration

	Face to Face	Face to Back
Communication Mémoire - Processeur	11.9 GHz	7.5 GHz
Communication Processeur - BGA	7.4 GHz	11.9 GHz

On note l'égalité entre les fréquences maximale de la communication Mémoire - Processeur en configuration Face to Face et de la communication Processeur - BGA en configuration Face to Back ainsi que celle entre la communication Mémoire - Processeur en configuration Face to Back et la communication Processeur - BGA en configuration Face to Face.

Ces égalités démontrent que les Cu-Pillar de diamètre 20  $\mu\text{m}$  et ceux de diamètre 55  $\mu\text{m}$  ont la même influence sur les fréquences maximales de communication. En effet, les briques élémentaires d'interconnexion constituant la communication Mémoire - Processeur (configuration Face to Face) sont presque identiques à celles constituant la communication Processeur - BGA (configuration Face to Back), la seule différence étant la nature des Cu-Pillar. Il en est de même en ce qui concerne les deux autres types de communications.

Ces résultats s'expliquent grâce aux valeurs résistives et capacitives des piliers de cuivre. En effet, l'impact de la résistance est minime, 7,4 m $\Omega$  en statique pour le Cu-Pillar-d20 et 2,1 m $\Omega$  en statique pour le Cu-Pillar-d55, par rapport à la valeur résistive du driver d'entrée (25  $\Omega$ ). Pour les capacités, bien que leurs géométries soient différentes, leurs capacités propres (prépondérantes devant les capacités de couplage) sont équivalentes, de l'ordre de 52 fF (cf. Chapitre 3 partie III.1.2).

On en conclut qu'avec des chaînes d'interconnexion réalisées sur des briques élémentaires d'interconnexion issues des technologies de moyenne densité d'intégration, les spécifications en termes de fréquence maximale de fonctionnement du Wide I/O sont atteintes, quelque soit l'empilement choisi (Face to Face ou Face to Back).



## II.2 L'interposer : une solution de plateforme technologique 3D prometteuse.

Ce sont les difficultés de routage sur le BGA, entraînant un surcoût non négligeable, qui incitent les concepteurs à penser à de nouvelles plateformes technologiques de type interposer. En effet, dans le but d'améliorer le rendement et la fiabilité des systèmes intégrés et de réduire les coûts de mise en boîtier, plusieurs puces peuvent être intégrées sur un substrat de silicium, nommé interposer.

C'est une technologie innovante qui permet de réduire les temps de mise sur le marché pour les produits CMOS de dernière technologie. Ainsi les interfaces hautes vitesses de type I/O et d'autres fonctions, difficiles à intégrer, comme le traitement des signaux analogiques/mixtes peuvent être co-intégrées sur un interposer avec une fabrication reposant sur un nœud technologique plus ancien mais très mature.

Les contraintes sur les interconnexions 3D présentes dans l'interposer sont rendues encore plus fortes dans le cas des connexions avec le BGA, car le chemin parcouru par le signal en provenance du processeur s'avère plus long. Les exigences sont fortes du point de vue fiabilité et les puces doivent durer plus de 10 ans, et ce dans des conditions environnementales souvent difficiles [Marchal, 2011].

Dans cette partie, les fréquences maximales de communication, Mémoire - Processeur et Processeur – BGA, vont être comparées avec celles obtenues sans interposer. L'intégrité des signaux transmis va également être vérifiée à l'aide de diagrammes de l'œil.

### II.2.1 Description de l'empilement pour une solution avec Interposer

Dans le cas d'un empilement avec un Interposer, la communication entre la mémoire (Wide I/O) et le processeur (SOC) s'effectue grâce à deux Cu-pillar de diamètre 20  $\mu\text{m}$  reliés sur le niveau supérieur de l'interposer par une interconnexion de RDL, d'épaisseur 10  $\mu\text{m}$ . La communication entre le processeur et le BGA s'effectue quant à elle par l'intermédiaire d'un premier Cu-Pillar de diamètre 20  $\mu\text{m}$ , situé entre le processeur et l'interposer, d'un TSV de diamètre 10  $\mu\text{m}$ , permettant de traverser l'interposer et d'un second Cu-Pillar de diamètre 55  $\mu\text{m}$ , faisant le lien avec le BGA (Figure 5-8).

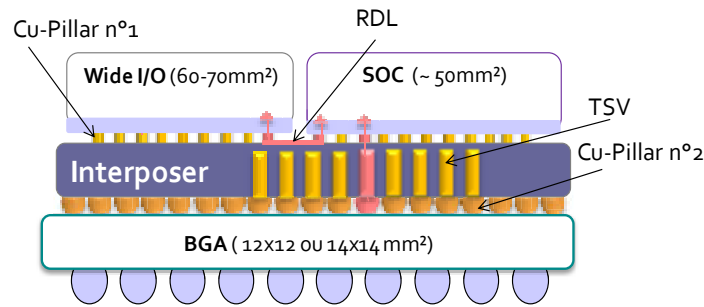


Figure 5-8 : Empilement avec interposer.

L'illustration ci-dessus, représente le cas le plus favorable, la mémoire et le processeur sont reliés par le chemin le plus court, soit une longueur de RDL d'environ 500  $\mu\text{m}$ . Ce scénario n'est cependant pas toujours possible. Afin de visualiser l'importance du placement des connexions entre deux puces (Mémoire - Processeur) la partie suivante expose les résultats pour trois configurations de routage, décrites sur la figure 5-9.

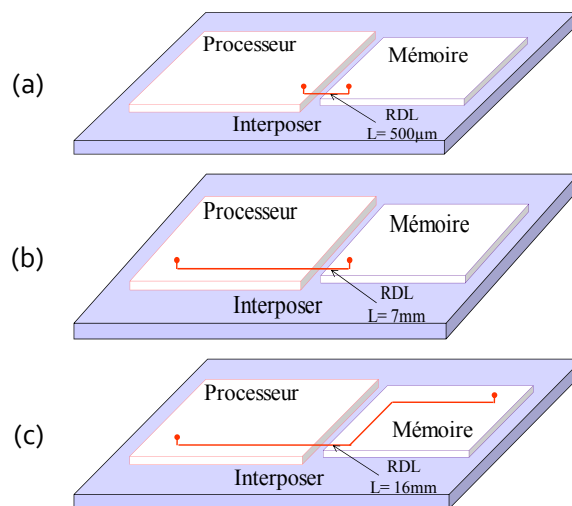


Figure 5-9 : Schémas de connexion avec interposer. (a) Cas idéal  $L_{\text{RDL}} = 500\mu\text{m}$ .  
 (b) Cas intermédiaire  $L_{\text{RDL}} = 7\text{ mm}$ . (c) Cas critique  $L_{\text{RDL}} = 16\text{ mm}$ .

La première (Figure 5-9-a) correspond au cas le plus favorable, explicité précédemment. La seconde configuration (Figure 5-9-b) correspond à une longueur d'interconnexion de RDL intermédiaire de l'ordre de 7 mm et la dernière (Figure 5-9-c) représente le cas extrême où les connexions sont à l'opposé sur chaque puce, soit environ 16 mm.

## II.2.2 Résultats obtenus sur les performances électriques

Cette partie donne les fréquences maximales de communication pour les trois configurations de routage, le retard engendré par chaque chaîne ainsi que l'intégrité des signaux transmis au travers différents diagrammes de l'œil.

### II.2.2.a Fréquences maximales des signaux

Le Tableau 5-2, synthétise les fréquences maximales de chaque chaîne pour les liaisons Mémoire-Processeur et pour la liaison processeur-BGA. En ce qui concerne la communication Processeur - BGA, la fréquence maximale de communication n'est que de 3,3 GHz ce qui ne satisfait pas le cahier des charges (5,8 GHz attendus) pour cette liaison. La valeur capacitive de 320 fF de la chaîne d'interconnexion complète est trop élevée pour permettre une transmission plus rapide. Dans ce cas une optimisation passe par la diminution des capacités de chaque brique élémentaire d'interconnexion ou par l'utilisation d'une densité d'intégration plus élevée. Pour les trois chaînes d'interconnexions représentant la communication Mémoire - Processeur, on remarque que les fréquences maximales varient énormément en fonction des longueurs d'interconnexion de RDL étudiées, 500  $\mu\text{m}$ , 7 mm et 16 mm. Elles donnent respectivement des fréquences maximales de 5,4 GHz, 1,16 GHz et 535 MHz.

Tableau 5-2 : Fréquences maximales des communications lors d'un empilement avec interposer.

Communication	Fréquence maximale
Mémoire – Processeur	( $L_{\text{RDL}} = 500 \mu\text{m}$ ) 5,4 GHz
	( $L_{\text{RDL}} = 7 \text{ mm}$ ) 1,16 GHz
	( $L_{\text{RDL}} = 16 \text{ mm}$ ) 535 MHz
Processeur – BGA	3,3 GHz

La Figure 5-10, représente la variation de la fréquence maximale de communication, en fonction de la longueur de l'interconnexion de RDL. Le fait qu'elle diminue très rapidement confirme l'importance du placement des connexions et donc de la conception des puces.

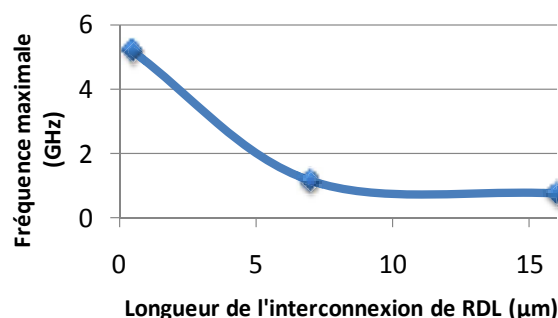


Figure 5-10 : Variation de la fréquence maximale en fonction de la longueur d'interconnexion de RDL.

Dans l'exemple de l'application Wide I/O, pour répondre au cahier des charges les communications Wide I/O – Processeur multimédia doivent supporter des fréquences de

transmission d'au moins 200 MHz et d'au moins 5,8 GHz, pour les communications Processeur multimédia – BGA. Avec un empilement comportant un interposer le Tableau 5-2 nous indique que les spécifications sont remplies quelle que soit la longueur de l'interconnexion de RDL pour la communication Wide I/O – Processeur multimédia, mais pas pour la communication Processeur multimédia – BGA. La réduction de la capacité parasite associée aux piliers de cuivre est indispensable pour atteindre la bande passante requise.

Cependant les fréquences maximales ne suffisent pas pour garantir une bonne transmission. Les signaux transmis le long d'une chaîne d'interconnexion utilisant les longueurs d'interconnexion supérieures au millimètre doivent être étudiés avec prudence, à cause des phénomènes de propagation dans les lignes de transmission. Les parties suivantes mettent l'accent sur les temps de retard à 50% et l'intégrité des signaux.

### II.2.2.b Étude du retard en fonction des longueurs de RDL.

Les temps de retard à 50% dépendent d'un grand nombre de paramètres. Dans nos études, les caractéristiques typiques de drivers de technologie 65 nm sont utilisées, avec une résistance d'accès de 25  $\Omega$  et une capacité de charge de 23 fF. Le but n'étant pas de déterminer avec exactitude le retard (trop dépendant de l'environnement) mais de dégager des tendances sur l'évolution des retards entre différentes chaînes d'interconnexion (Tableau 5-3).

Tableau 5-3 : Retards à 50% pour des liaisons Mémoire – Processeur avec interposer

Longueur de l'interconnexion de RDL	Retard à 50 %
500 $\mu\text{m}$	4,9 ps
7 mm	51 ps
16 mm	132,9 ps

D'après les résultats ci-dessus, utiliser un routage avec une longueur d'interconnexion de RDL de 7 mm ou de 16 mm, au lieu d'un routage optimisé ( $L_{\text{RDL}} = 500 \mu\text{m}$ ), augmente le retard à 50% des signaux transmis de façon considérable.

La conclusion est flagrante, même si utiliser des longueurs d'interconnexions de 7 et 16 mm continue à remplir le cahier de charges d'une application de type Wide I/O, l'augmentation des retards met en évidence l'importance de la position des contacts. Pour appuyer ces propos, les diagrammes de l'œil des signaux transmis sur les des trois chaînes d'interconnexion pouvant faire le lien entre la mémoire et le processeur ont été réalisés.

### II.2.2.c Etude de l'intégrité des signaux à l'aide de diagramme de l'œil.

En effet le diagramme de l'œil permet, de manière très simple, d'apprécier et d'estimer la qualité des signaux numériques reçus. Il s'obtient en superposant les signaux de données reçus et en synchronisant le balayage au rythme de la durée des symboles. Les signaux reçus sont issus de la convolution de la réponse impulsionnelle d'un système avec des séquences aléatoires binaires (Figure 5-11). Dans les études suivantes les séquences aléatoires sont au nombre de 80 et contiennent chacune 30 symboles.

La Figure 5-12, représente les diagrammes de l'œil des signaux transmis sur les chaînes d'interconnexion pouvant faire le lien entre la mémoire et le processeur, pour deux débits binaires (5 et 10 GBy/s). Le débit binaire correspond à l'inverse de la durée d'un symbole et la durée d'un symbole est égale au nombre de point par symbole multiplié par la résolution temporelle.

Pour les études suivantes nous avons choisi :

- Résolution temporelle = 2 ps.
- Nombre de point par symbole = 50 (pour un débit binaire de 10 GBy/s) et 100 (pour un débit binaire de 5 GBy/s).

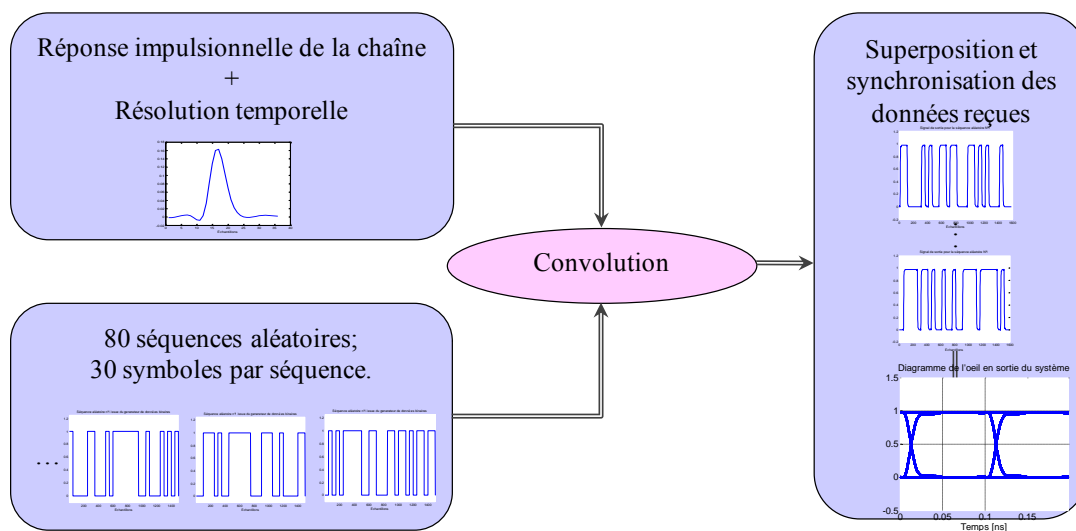


Figure 5-11 : Principe de réalisation d'un diagramme de l'œil.

Pour une longueur d'interconnexion de RDL de 500  $\mu\text{m}$ , la Figure 5-12 montre qu'en plus de satisfaire les fréquences maximales, l'intégrité du signal pour des débits binaires de 5 et 10 Gby/s, est conservée. En effet pour ces deux débits binaires l'ouverture de l'œil est égale à 97,6 %. Dans cette configuration le signal n'est pas dégradé.

Pour une longueur de 7 mm, une déformation du signal commence à apparaître. Phénomène accentué par l'augmentation du débit binaire. On remarque que pour un débit binaire de 5 Gby/s

l'ouverture de l'œil est de 84 % alors que pour 10 Gby/s l'ouverture de l'œil est de 65 %. Ceci est majoritairement dû au fait que les séquences aléatoires, de plus en plus agressives, sont mal gérées par des interconnexions si longues. Cette tendance se confirme lorsque l'on observe les diagrammes de l'œil effectués pour une interconnexion de 16 mm. En effet pour un débit binaire de 5 Gby/s l'ouverture de l'œil est de 59 % et de 36 % pour 10 Gby/s.

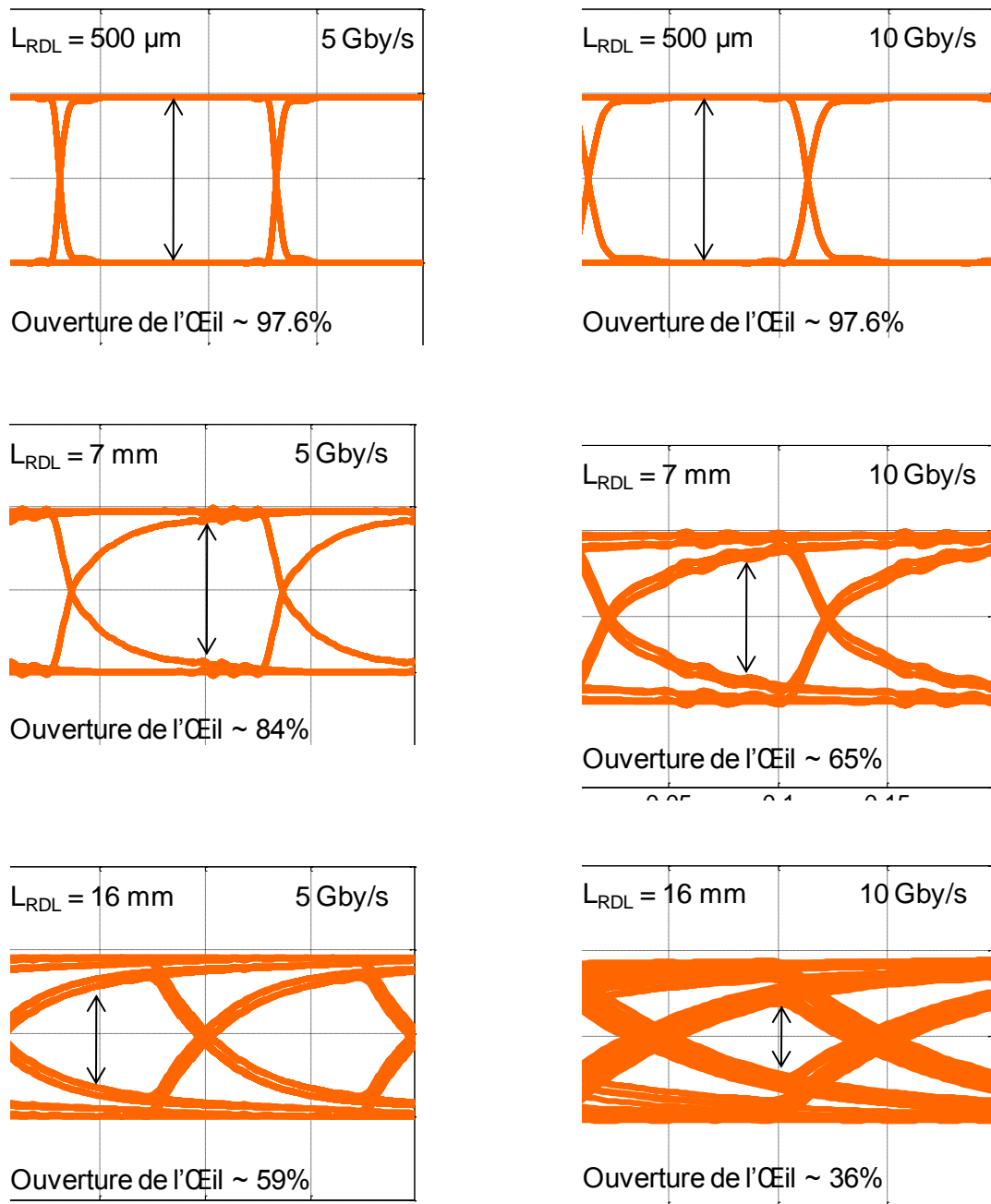


Figure 5-12 : Diagramme de l'œil correspondant aux communications Mémoire-Processeur dans le cas d'une solution avec interposer, pour des débits binaires de 5 et 10 Gbps.

Lorsque le signal est dégradé (faible ouverture de l'œil), une mauvaise synchronisation sur le signal de sortie engendre une fausse interprétation des niveaux logiques.

En conclusion la configuration la plus favorable ( $L_{RDL} = 500\mu\text{m}$ ) n'engendre aucun problème de transmission du signal pour la communication Mémoire - Processeur. Dans le cas intermédiaire, l'intégrité des signaux visualisés par le diagramme de l'œil est satisfaisante, cependant le retard augmente de manière très importante par rapport à une communication optimisée ( $L_{RDL} = 500\mu\text{m}$ ). Il ne faut donc pas négliger l'importance de la position des contacts. Pour finir le cas extrême est à proscrire ( $L_{RDL} = 16\text{ mm}$ ) même si le critère de fréquence maximale reste correct pour un certain nombre d'applications, le retard engendré est vraiment trop important et les signaux sont trop dégradés.

L'utilisation d'un interposer n'apporte pas plus de rapidité et ne respecte pas le cahier des charges en termes de fréquence maximale pour une liaison Processeur multimédia - BGA lors d'une application de type Wide I/O. Cependant lorsque les longueurs d'interconnexions de RDL sont relativement courtes la liaison Mémoire - Processeur multimédia répond amplement au cahier des charges. Les principaux avantages de l'approche interposer sont la réduction des délais de mise sur le marché (réutilisation de circuits existants), la possibilité de réaliser des systèmes complets sur silicium, la simplification du routage dans le BGA, qui constitue aujourd'hui un frein important à la densification des circuits et à la réduction des coûts de fabrication, ou encore la possibilité de co-intégrer des composants initialement difficiles à intégrer (composants passifs, MEMS, ...).

### II.3 Conclusion des parties 2 et 3.

Les parties précédentes montrent que pour un empilement sans interposer les liaisons Mémoire – Processeur et Processeur - BGA, réalisées avec des TSV de diamètre  $10\mu\text{m}$ , des Cu-pillar de diamètre  $20\mu\text{m}$  (utilisés entre le processeur et la mémoire) et les Cu-pillar de diamètre  $55\mu\text{m}$  (utilisés entre le processeur et le BGA), présentent des fréquences maximale qui varient entre 7,4 GHz et 11,9 GHz, satisfaisant les spécifications d'une application Wide I/O.

En ce qui concerne l'empilement avec interposer le cahier des charges la communication Processeur – BGA n'est pas respecté, la fréquence maximale de communication n'est que de 3,3 GHz au lieu des 5,8 GHz attendus.

Cependant ces études ont été menées avec les briques élémentaires issues de technologies de moyenne densité d'intégration. Avec des briques élémentaires de densité d'intégration différentes, les dimensions des éléments constituant les chaînes d'interconnexions vont varier et de fait les fréquences maximales.

Les configurations d'empilement sans interposer, en Face to Face et en Face to Back et avec interposer sont repris dans la partie suivante pour diverses densités d'intégration. Tout le processus

d'étude sera détaillé dans un premier temps pour l'empilement Face to Back, puis les conclusions sur les fréquences maximales des trois empilements étudiés en fonction de quatre densités d'intégration sont exposées et pour finir une stratégie de design pour chaque empilement est déterminée.

### III Communications Mémoire-Processeur et Processeur-BGA en fonction des densités d'intégration (avec et sans interposer).

Il a été défini que l'utilisation de briques élémentaires d'interconnexion issues de technologies de moyenne densité d'intégration, avec des TSV de diamètre 10  $\mu\text{m}$  et des Cu-pillar de diamètre 55  $\mu\text{m}$ , permettait d'atteindre les spécifications de la plupart des liaisons hormis la liaison Processeur-BGA de l'empilement avec interposer. Mais qu'en est-il de ces fréquences maximales, lorsque que le circuit est intégré avec d'autre densités d'intégration ? Dans cette partie, les fréquences maximales, pour les trois configurations d'empilement, sont déterminées en fonction de plusieurs densités d'intégration.

Dans un premier temps, les communications Mémoire - Processeur et Processeur - BGA d'une structure à empilement Face to Back sont étudiées en détail pour cinq densités d'intégration. Par la suite les fréquences maximales des trois empilements en fonction des densités d'intégration sont exposées sous forme d'abaque et pour finir une stratégie pour chaque empilement est déterminée

#### III.1 Cas de l'empilement Face to Back

##### III.1.1 Description de l'empilement

L'empilement de trois niveaux est constitué d'un BGA en puce inférieure, d'un processeur en puce centrale et d'une mémoire en puce supérieure. La communication entre la mémoire et le processeur est effectué par un premier Cu-Pillar, un TSV et une interconnexion de BEOL de longueur 20  $\mu\text{m}$ . Le processeur et le BGA sont quant à eux, reliés par un second Cu-Pillar, comme le montre la figure 5-13.

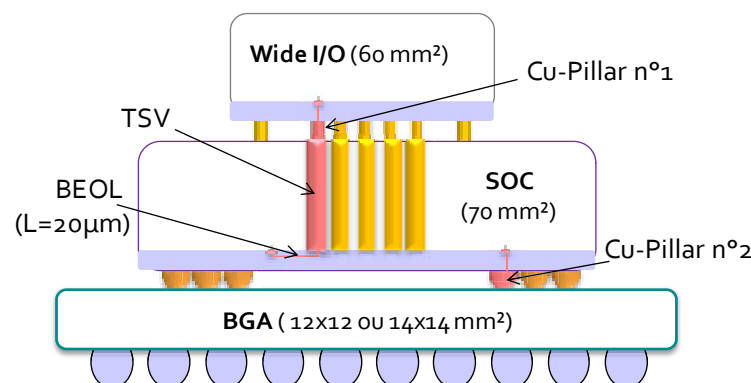



Figure 5-13 : Eléments d'interconnexion constituant les communications inter-niveaux.



Afin de visualiser les performances des deux communications (mémoire-processeur SOC et processeur SOC-BGA) en fonction de la densité d'intégration, cinq types de chaîne d'interconnexion sont étudiées.

La liste des éléments d'interconnexion, variant en fonction des densités d'intégration dans chaque chaîne, est présentée dans le tableau 5-4. L'interconnexion de BEOL et le premier Cu-Pillar de diamètre 20  $\mu\text{m}$  sont les mêmes dans les cinq chaînes.

Tableau 5-4 : Briques d'interconnexion utilisées en fonction des densités d'intégration.

		Wide I/O - SOC	SOC - BGA
Densité d'intégration  Faible Forte	Chaîne 1	TSV-d60	Cu-Pillar-d75
	Chaîne 2	TSV-d10	Cu-Pillar-d55
	Chaîne 3	TSV-d6	Cu-Pillar-d55
	Chaîne 4	TSV-d5	Cu-Pillar-d55
	Chaîne 5	TSV-d3	Cu-Pillar-d35

La première chaîne, nommée chaîne 1, utilisant un TSV de diamètre 60  $\mu\text{m}$  et un second Cu-Pillar de diamètre 75  $\mu\text{m}$ , est associée à de la basse densité d'intégration. Les trois chaînes suivantes, notées chaîne 2, 3 et 4, utilisant respectivement des TSV de diamètres 10, 6 et 5  $\mu\text{m}$  ainsi que des second Cu-Pillar de diamètre 55  $\mu\text{m}$ , sont assimilées à de la moyenne densité d'intégration. Pour finir, la dernière chaîne notée chaîne 5, appartenant à de la haute densité d'intégration, utilise un TSV de diamètre 3  $\mu\text{m}$  et un second Cu-Pillar de diamètre 35  $\mu\text{m}$ .

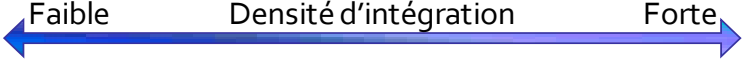
### III.1.2 Résultats sur l'empilement Face to Back.

Cette partie expose les fréquences maximales de communication, pour une configuration Face to Back, ainsi que l'intégrité des signaux transmis au travers de différents diagrammes de l'œil. Pour finir une « Roadmap » en termes d'évolution de performance est déterminée pour cet empilement.

#### III.1.2.a Fréquence des signaux

Le Tableau 5-5, synthétise les fréquences maximales de chaque chaîne selon leur densité d'intégration. Pour les chaînes d'interconnexion servant à une communication Mémoire – Processeur et à une communication Processeur - BGA, on remarque que les fréquences maximales sont assez élevées, et ceci pour les cinq densités d'intégration. Elles varient entre 4,7 GHz et 9,6 GHz pour la communication Mémoire – Processeur et entre 8,3 GHz et 15,2 GHz pour la communication Processeur - BGA. Ce qui remplit totalement les spécifications attendues pour les performances mémoires de type LPDDR2, LPDDR3 ou encore Wide I/O.

Tableau 5-5 : Fréquences maximales en fonction des densités d'intégration.

					
	Chaîne 1	Chaîne 2	Chaîne 3	Chaîne 4	Chaîne 5
Communication	4.7 GHz	7.5 GHz	8.6 GHz	8.9 GHz	9.6 GHz
Mémoire - Processeur					
Communication	8.3 GHz	11.9 GHz	11.9 GHz	11.9 GHz	15.2 GHz
Processeur - BGA					

Cet empilement est optimal, il favorise la vitesse de communication Processeur – BGA, sur laquelle il y a le plus de contraintes, et permet des fréquences maximales de communication suffisamment élevées pour satisfaire les spécifications quelle que soit la densité d'intégration utilisée.

### III.1.2.b Intégrité des signaux - diagrammes de l'œil.

Afin de s'assurer de la conformité des signaux transmis en haute densité d'intégration, les diagrammes de l'œil des signaux sur les liaisons sont tracés en Figure 5-14, pour deux débits binaires de 5 et 10 GBy/s. On remarque de très bonnes performances, pour les deux communications (Mémoire - Processeur et Processeur - BGA), rendant possible toute application de type mémoire sur processeur.

Pour ces deux communications le constat est le même. Les signaux ne sont pas déformés et les deux diagrammes de l'œil ont des ouvertures très bonnes, de l'ordre de 98 % pour la transmission Mémoire - Processeur et de l'ordre de 100 % pour la communication Processeur - BGA, et ce pour les deux débits binaires de 5 et 10 GHz.

Ce constat conforte celui obtenu par le critère de fréquence maximale. De plus, les hautes densités d'intégration donnent les meilleurs performances, ainsi la loi de Moore «Smaller is Better » se vérifie également pour les interconnexions 3D.



Figure 5-14 : Diagrammes de l'œil des communications Mémoire-Processeur (à gauche) et Processeur-BGA (à droite) dans le cas d'une technologie à haute densité d'intégration.

A la vue des très bonnes performances visualisées sur les diagrammes de l'œil, ce sont les contraintes en terme de fréquence maximale, qui vont permettre de déterminer la densité d'intégration la plus adaptée en fonction de l'application visée.

On peut également noter que les Cu-Pillar nécessaires à une communication Processeur – BGA, pour des générations technologiques utilisant des TSV de diamètre 10 et 6  $\mu\text{m}$  sont identiques. Ce sont des Cu-Pillar de diamètre 55  $\mu\text{m}$  et de hauteur 80  $\mu\text{m}$ . Cette hauteur est fixée par le procédé de fabrication utilisant un système de remplissage par capillarité de l'underfill entre les Cu-Pillar. Actuellement, est en court de développement, une technologie qui permettrait de coller un processeur à un BGA avec un système de thermo compression. Cette technique permettrait l'utilisation d'un Cu-Pillar de diamètre 55  $\mu\text{m}$  et de hauteur 40  $\mu\text{m}$ . Cette brique est développée indépendamment du 3D pour les technologies de report par flip-chip, qui remplacent petit à petit les bump comme alternative au wire-bonding (câblage filaire dans le boîtier)


Cette nouvelle brique élémentaire, associée dans la partie suivante à la génération technologique utilisant des TSV de diamètre 6  $\mu\text{m}$ , va optimiser les performances en fréquence maximale de fonctionnement de la communication Processeur-BGA de cette génération.

Une « RoadMap » en termes d'évolution de performances tend à se dessiner concernant l'intégration 3D. En effet il est possible de distinguer quatre catégories de TSV, représentant chacune une génération technologique. Le passage d'une génération à une autre entraîne une forte augmentation du nombre de TSV par  $\text{mm}^2$  et une amélioration non négligeable des vitesses de communication entre une mémoire et un processeur et entre un processeur et un BGA.

Le Tableau 5-6 donne la densité de TSV par millimètre carré et l'évolution des performances d'une génération à une autre. On remarque qu'avec un empilement Face to Back toutes les générations permettent de satisfaire les spécifications en termes de fréquences maximales de communication du Wide I/O, soit de 200MHz pour la communication Mémoire – Processeur et 5,8GHz pour la communication Processeur – BGA.

Cependant, le passage d'une génération technologique à une autre offre de très bonnes perspectives d'évolution. On remarque, par exemple, que le passage d'une faible à une moyenne densité d'intégration augmente d'environ 60 % la vitesse de communication entre une mémoire et un processeur.

Tableau 5-6 : RoadMap de « l'intégration 3D » pour un empilement Face to Back.

				
Génération technologique (diamètre des TSV)	60 $\mu$	10 $\mu$	6 $\mu$	3 $\mu$
Densité (TSV/mm <sup>2</sup> )	16	400	441	5776
Vitesse				
Mémoire - Processeur (GHz)	4,7	7,5	8,6	9,6
Amélioration par génération (%)		59,5	14,6	11,6
Vitesse				
Processeur - BGA (GHz)	8,3	11,9	12	15,2
Amélioration par génération (%)		43,3	0,8	26,6

### III.2 Roadmap et stratégie d'empilement pour les circuits associant une mémoire à un processeur.

Cette partie s'intéresse aux performances, en termes de fréquences maximales de communication, entre une mémoire et un processeur et entre un processeur et un BGA, en fonction des densités d'intégration. Tout d'abord, un abaque donnant les fréquences maximales, des communications Mémoire – Processeur et Processeur – BGA, pour des empilements avec et sans interposer, en fonction des densités d'intégration est présenté. Puis les stratégies d'empilement à adopter pour chaque type de niveau d'intégration sont déterminées.

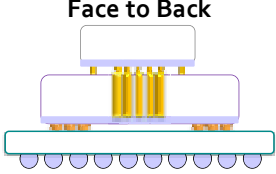
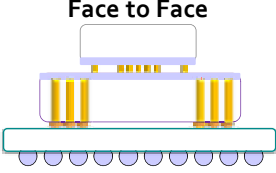
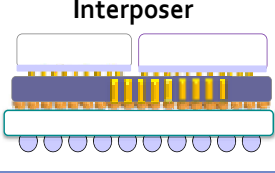
#### III.2.1 Fréquences maximales en fonction des générations technologiques.

Il a été démontré dans la partie précédente que l'empilement Face to Back est optimal. Il favorise la communication présentant le plus de contraintes (Processeur – BGA). Avec cette configuration le cahier des charges d'une application Wide I/O est totalement rempli.

Cependant les autres empilements, utilisant certaines densités d'intégrations, faible densités d'intégration pour l'empilement Face to Face et faible à moyenne densités d'intégration pour

l'empilement avec interposer, présentent des limites en fréquence trop basses pour remplir les spécifications requises (Tableau 5-7).

Tableau 5-7 : Abaque des fréquences maximales en fonction des générations technologiques.

		<div style="display: flex; align-items: center; justify-content: center;"> <span style="margin-right: 10px;">← Faible</span> <span>Densité d'intégration</span> <span style="margin-left: 10px;">→ Forte</span> </div>			
Génération technologique (Diamètre des TSV)		60 $\mu$	10 $\mu$	6 $\mu$	3 $\mu$
<b>Face to Back</b> 	Mémoire-Processeur (GHz)	4,7	7,5	8,6	9,6
	Processeur-BGA (GHz)	8,3	11,9	12	15,2
<b>Face to Face</b> 	Mémoire-Processeur (GHz)	11,9	11,9	11,9	11,9
	Processeur-BGA (GHz)	3,8	7,4	8,6	15
<b>Interposer</b> 	Mémoire-Processeur (GHz)	5,4	5,4	5,4	5,4
	Processeur-BGA (GHz)	0,711	3,3	3,5	8,1

L'empilement Face to Face offre majoritairement de bonnes performances, hormis pour la plus faible densité d'intégration qui a, elle, une fréquence maximale inférieure à 5,8 GHz pour la communication Processeur – BGA.

C'est dans le cas d'un empilement utilisant un interposer que le constat est le plus mitigé. Même si les spécifications d'une application Wide I/O sont atteintes pour la communication Mémoire – Processeur quelle que soit la densité d'intégration, pour la communication Processeur – BGA, seule la plus forte densité d'intégration utilisant des TSV de diamètre 3  $\mu$ m, offre une fréquence maximale supérieure à 5,8 GHz. Dans ce cas l'utilisation d'un interposer n'est valable que pour de fortes densités d'intégration.

La partie suivante donne la stratégie d'intégration la mieux adaptée pour chaque empilement.

### III.2.2 Stratégies d'empilement selon le type de densité d'intégration.

Un empilement Face to Back est optimum. Les fréquences de communications remplissent le cahier des charges d'une application Wide I/O quelle que soit la densité d'intégration.

Dans le cas d'un empilement Face to Face il faut privilégier les moyennes et forte densités d'intégration pour couvrir les spécifications demandées. Pour de faibles densités d'intégration la

fréquence maximale d'une communication Processeur-BGA ne dépasse pas les 3,8 GHz alors que la spécification impose 5,8 GHz

Pour finir l'empilement avec interposer doit opter pour de fortes densités d'intégration avec de faibles longueurs de RDL entre la mémoire et le processeur. Lors de l'utilisation de faible ou de moyenne densité d'intégration, les contraintes du cahier des charges ne sont pas satisfaites.

## IV Conclusion

Ce chapitre donne les spécifications en termes de fréquence maximale de communication et d'intégrité des signaux, pour les différents empilements permettant l'association d'une mémoire et d'un processeur, pour plusieurs densités d'intégration. L'exemple d'une mémoire Wide I/O a été pris mais les résultats obtenus sont applicables à tous types d'application présentant un empilement similaire.

Dans un premier temps les études ont été menées sur une densité d'intégration bien particulière afin de présenter la démarche suivie. Puis nos recherches se sont étendues à de nombreuses densités d'intégration et à divers types d'empilement.

Nous avons distingué quatre catégories de TSV, représentant chacune une génération technologique. Une « Roadmap » en termes d'évolutions des performances a été établie pour l'empilement Face to Back, mettant en évidence les améliorations provoquées par le passage d'une génération technologique à une autre. Le nombre de TSV par mm<sup>2</sup> ainsi que les vitesses de communication entre une mémoire et un processeur et entre un processeur et un BGA, s'en trouvent considérablement améliorés. On remarque, par exemple, que le passage d'une basse à une moyenne densité d'intégration augmente d'environ 60 % la vitesse de communication entre une mémoire et un processeur et d'environ 43% la vitesse de communication entre un processeur et un BGA.

Par la suite un « abaque » définissant les vitesses de communication, entre une mémoire et un processeur et entre un processeur et un BGA, utilisable pour tout type d'application, en fonction de différents empilements a été défini. On a, entre autre, mis en évidence que l'empilement avec interposer est celui qui présente le plus de contraintes. En effet seule une haute densité d'intégration, donnant une fréquence de 8,1 GHz, permet de satisfaire les spécifications requises, de 5,8 GHz, d'une communication Processeur - BGA.

Pour finir, une stratégie pour chaque empilement a été établie. L'empilement Face to Back est la configuration optimale qui respecte le cahier des charges d'une application Wide I/O, quelle que soit la densité d'intégration.

Lors d'un empilement Face to Face, il faut privilégier les moyennes et fortes densités d'intégration et pour un empilement avec interposer il faut opter pour des fortes densités d'intégration.

Dorénavant l'application Wide I/O n'est plus obligatoirement réalisée en Face to Back, en effet cet empilement est, pour le moment, privilégié par les concepteurs car pressenti comme celui donnant les meilleures performances. Il est possible, en utilisant les bonnes densités d'intégration, d'utiliser des empilements de type Face to Face si les contraintes d'empilements l'imposent ou une plateforme technologique telle qu'un interposer si il devient nécessaire de réduire les coûts ou les délais de mise sur le marché.





# Conclusion générale et perspectives

---

Ces travaux de thèse portaient sur la caractérisation, la modélisation et l'optimisation des performances électriques des réseaux d'interconnexions et des composants passifs dans les empilements 3D de circuits intégrés.

Les objectifs de ces travaux étaient :

- La caractérisation expérimentale et la modélisation de chacune des briques élémentaires d'interconnexion présente dans un réseau d'interconnexions d'un empilement 3D, et cela sur des spectres de fréquences conformes à ceux des signaux véhiculés dans les différentes applications visées.
- Le développement des outils de simulation des performances électriques des réseaux d'interconnexions 3D, réseaux réalisés en assemblant les différentes briques élémentaires d'interconnexions spécifiques aux technologies d'intégration 3D.
- L'analyse et la prédiction des performances électriques de ces réseaux d'interconnexions 3D en fonction des différentes alternatives technologiques d'empilement et niveaux d'intégration.
- L'extraction de règles de conception, c'est-à-dire les meilleures stratégies de routage, d'orientation des puces et de densité d'intégration permettant d'optimiser les performances des réseaux d'interconnexions dans les empilements 3D.

Ces travaux se terminent par une mise en œuvre pratique de nos résultats : l'évaluation et l'optimisation des performances des réseaux d'interconnexion spécifiques à une application concrète, l'intégration 3D de circuits mémoire Wide I/O et processeur Multimédia.

## Bilan

Dans la première partie du second chapitre, l'importance des techniques d'épluchage permettant d'extraire les caractéristiques intrinsèques des dispositifs sous test enfouis dans un environnement complexe, a été mise en avant. Nous avons présenté une méthode de de-embedding générique applicable à tous types de dispositifs sous test ainsi que quatre méthodes de de-embedding bien

adaptées à l'extraction de l'impédance caractéristique d'interconnexions enfouies dans un circuit intégré. Dans la seconde partie nous avons présenté les résultats de caractérisation obtenus avec ces méthodes de de-embedding sur les différentes briques élémentaires d'interconnexion composant une chaîne d'interconnexion complète dans un empilement 3D. Les interconnexions de RDL, les interconnexions de BEOL, les vias traversant le silicium et les piliers de cuivre ont pu être étudiés et les modèles extraits comparés à ceux obtenus par simulation électromagnétique. Un composant passif réactif a également été caractérisé et nous avons mis en évidence l'utilité d'intégrer des inductances sur les couches de redistribution afin de maintenir de bons facteurs de qualité.

Les bons accords obtenus entre résultats expérimentaux et résultats de calcul ont permis de conclure quant à la fiabilité des outils de caractérisation et de modélisation.

Dans le chapitre trois, nous avons étudié toutes les briques élémentaires pouvant constituer une chaîne d'interconnexions en intégration 3D, afin d'obtenir les performances de chaque élément (interconnexion de RDL, TSV, interconnexion de BEOL, Cu-Pillar et inductance) en termes de retard ou de facteur de qualité.

Dans le cas des interconnexions de RDL, la performance des lignes à haute densité, intégrées dans du BCB, a été démontrée. Pour les interconnexions du BEOL, le retard engendré dépend principalement de la longueur de l'interconnexion. Elle doit être inférieure à 1,4 mm afin de concurrencer les performances des interconnexions de RDL. Le retard induit par chaque TSV dépend de sa densité d'intégration : un TSV intégré à haute densité induit beaucoup moins de retard qu'un TSV intégré à faible densité. Concernant les Cu-Pillars, il a été observé que le retard induit est négligeable par rapport au retard engendré par les autres briques d'interconnexion.

Un modèle complet d'une inductance planaire deux ports a également été exposé afin d'exprimer les performances des inductances insérées dans une architecture 3D.

De plus, deux études paramétriques poussées, sur l'impact d'une architecture 3D sur les performances de dispositifs déjà présents en intégration 2D (les interconnexions horizontales et les inductances), ont permis de définir plusieurs règles d'optimisation lors d'une intégration 3D :

Pour les interconnexions horizontales :

- 1) Dans une configuration Face to Back, l'épaisseur de collage doit être deux fois plus élevée que l'espacement ligne-masse pour une interconnexion de type CPW. Dans le cas contraire elle doit être maximisée.
  - ↳ Passer d'une épaisseur de collage de 100 nm à une épaisseur de 200 nm, avec un espacement entre les métallisations de 800 nm permet de réduire le retard de 17 %.

- 2) L'espacement ligne-masse des interconnexions CPW doit être maximisé.
  - ↘ Passer d'un espacement entre les métallisations de 800 nm à 1200 nm, pour une épaisseur de collage de 200 nm permet de réduire le retard de 11 %
- 3) L'architecture des interconnexions doit être adaptée aux contraintes de fabrication.
  - ↘ Pour une épaisseur de collage de 100 nm, privilégier une topologie MSW permet de réduire le retard de 34 %
- 4) Un compromis doit être trouvé entre l'optimisation du retard et de la diaphonie.
  - ↘ Passer d'une épaisseur de collage de 100 nm à une épaisseur de 200 nm, avec un espacement entre les métallisations de 800 nm permet de réduire le retard de 25 % (lors d'une excitation seule) mais augmente le niveau de diaphonie de 3,5 %.
  - Passer d'un espacement entre les métallisations de 800 nm à 1200 nm, pour une épaisseur de collage de 1  $\mu$ m permet de réduire le retard de 12 % (lors d'une excitation seule) mais augmente le niveau de diaphonie de 2,2%.
- 5) Les dummies doivent être éloignés de la ligne signal, soit grâce à l'épaisseur de collage, soit par une zone d'exclusion.
  - ↘ Avec une épaisseur de collage de 100 nm et pour un dummy qui passe du dessus de l'interconnexion à une distance de 800 nm, le retard diminue de 48 %.
- 6) Dans une configuration Face to Face, l'épaisseur de collage doit être trois fois plus élevée que l'espacement ligne-masse. Dans le cas contraire elle doit être maximisée.
  - ↘ Avec un dummy situé juste au dessus de l'interconnexion et pour une épaisseur de collage passant de 100 nm à 1  $\mu$ m, le retard diminue de 55 %.

Pour les inductances, en configuration Face to Back:

- 1) Le substrat de silicium rapporté doit être le plus mince possible.
  - ↘ Amincir le substrat de la puce rapportée de 100  $\mu$ m à 5  $\mu$ m permet d'augmenter l'amplitude de facteur de qualité de 8,5 %.
- 2) La conductivité du silicium doit être la plus faible possible.
  - ↘ Opter pour un substrat ayant une conductivité de 2 S/m plutôt qu'un substrat ayant une conductivité de 50 S/m permet d'augmenter l'amplitude du facteur de qualité de 15 %.
- 3) Il faut privilégier les fortes épaisseurs de collage.
  - ↘ Par exemple, une épaisseur de diélectrique entre l'inductance et le substrat de la puce supérieure, de 4  $\mu$ m plutôt que de 1  $\mu$ m, permet d'augmenter l'amplitude du facteur de qualité de 13 %.

Dans le quatrième chapitre, l'association de ces briques élémentaires en fonction de différentes stratégies d'intégration donnent quelques règles majeures à suivre, que ce soit en termes de routage, d'orientation ou encore d'option technologique.

Les interconnexions de RDL offrent de meilleures performances de routage par rapport à une interconnexion de BEOL pour des distances supérieures à 1,4 mm grâce à l'isolation low-k et à la forte épaisseur de métallisation. Plus les longueurs d'interconnexion nécessaires au routage sont élevées, plus il est rentable d'utiliser des interconnexions de RDL plutôt que des interconnexions de BEOL. Une diminution de 60 % sur ce retard peut être atteinte pour une longueur totale d'interconnexion de 2 mm.

L'orientation de la puce centrale, lors d'un empilement à trois niveaux doit résulter d'une analyse claire des spécifications du produit afin d'optimiser au mieux une communication puce à puce ou puce à substrat, en fonction d'une configuration Face to Face ou Face to Back. Pour une optimisation de la rapidité de communication puce à puce, la meilleure configuration est le Face to Back. Par contre, si l'optimisation doit se faire sur la communication puce à substrat alors il faudra privilégier une configuration Face to Face. Une diminution de 75 % sur le retard peut être atteinte pour des liaisons longues (5 mm).

L'élément critique dans une chaîne d'interconnexions dépend principalement de la densité d'intégration. La loi de Moore est également valable en intégration 3D. Ces recommandations générales constituent une première étape pour aider les concepteurs à trouver les meilleurs scénarios d'intégration 3D pour leurs produits. Plus la densité d'intégration augmente, plus rapidement l'interconnexion de RDL devient l'élément critique, et ce d'autant plus pour des longueurs élevées. Cependant pour des interconnexions de RDL relativement courtes, inférieures à 1,4 mm, pour de faibles densités d'intégration ou inférieures à 400  $\mu\text{m}$  pour les moyennes densités d'intégration c'est le TSV qui est l'élément à optimiser.

Ces recommandations générales constituent une première étape pour aider les concepteurs à trouver les meilleurs scénarios d'intégration 3D pour leurs produits.

Tous ces éléments nous permettent de mieux appréhender les performances possibles, d'applications réalistes, en fonction de leurs designs, de leurs empilements ou encore de leurs densités d'intégration.

Enfin, le dernier chapitre est consacré à une application prometteuse de l'intégration 3D. Il traite des possibilités offertes par l'intégration 3D pour atteindre les spécifications requises en termes de fréquence maximale de communication et d'intégrité des signaux, pour les différents empilements

permettant l'association d'une mémoire et d'un processeur, et cela pour plusieurs densités d'intégration. L'exemple d'une mémoire Wide I/O a été pris mais les résultats obtenus sont applicables à tous types d'applications présentant un empilement similaire.

Dans un premier temps les études ont été menées pour une densité d'intégration bien particulière, la moyenne densité, afin de présenter notre démarche d'analyse. Puis nos recherches se sont étendues aux différents niveaux de densités d'intégration et à différents types d'empilements.

Nous avons distingué quatre catégories de TSV, représentant chacune une génération technologique. Une « Roadmap » en termes d'évolution des performances a été établie pour l'empilement Face to Back, mettant en évidence les améliorations provoquées par le passage d'une génération technologique à une autre. Le nombre de TSV par mm<sup>2</sup> ainsi que les vitesses de communication entre une mémoire et un processeur et entre un processeur et un BGA, s'en trouvent considérablement améliorés. On remarque, par exemple, que le passage d'une faible à une moyenne densité d'intégration augmente d'environ 60 % la vitesse de communication entre une mémoire et un processeur et d'environ 43% la vitesse de communication entre un processeur et un BGA.

Par la suite un « abaque » définissant les vitesses de communication, entre une mémoire et un processeur et entre un processeur et un BGA, utilisable pour tout type d'application, en fonction de différents empilements a été proposé. On a entre autre mis en évidence que l'empilement avec interposer est celui qui présente le plus de contraintes. En effet seule une haute densité d'intégration, donnant une fréquence maximale de 8,1 GHz, lui permet de satisfaire les spécifications requises, de 5,1 GHz, d'une communication Processeur-BGA.

Pour finir, une stratégie d'empilement a été déterminée pour le cas du Wide I/O en fonction des densités d'intégration.

L'empilement Face to Back est la configuration optimale qui satisfait le cahier des charges d'une application Wide I/O quelque soit la densité d'intégration.

Lors d'un empilement Face to Face, il faut privilégier les moyennes et fortes densités d'intégration et pour un empilement avec interposer il faut opter pour de fortes densités d'intégration.

Dorénavant l'application Wide I/O n'est plus obligatoirement réalisée en Face to Back, en effet cette empilement est, pour le moment, privilégié par les concepteurs car pressenti comme celui donnant les meilleures performances.

Cependant il est maintenant possible, en utilisant les densités d'intégration adéquates, d'utiliser des empilements de type Face to Face si les contraintes d'empilements l'imposent ou une

plateforme technologique tel qu'un interposer s'il s'avère nécessaire de réduire les coûts ou les délais de mise sur le marché.

## **Perspectives**

Tout d'abord, dans un esprit d'auto-évaluation, les travaux que nous avons exposés peuvent être complétés sur plusieurs aspects. En particulier nous pouvons mentionner deux types d'études à mener :

- Elargissement des résultats à plusieurs autres critères de performance. En effet les critères de performance sur lesquels nous nous sommes focalisés ont été le temps de propagation à 50 % le long des différents réseaux d'interconnexions que nous avons expertisés et les diagrammes de l'œil associés. Ce choix a été justifié par le fait que ces critères sont très représentatifs de l'intégrité des signaux transmis. Néanmoins, selon les applications visées, il serait intéressant d'appuyer nos études sur des analyses de performance en termes :
  - de temps de montée des signaux (temps de passage de 10 % à 90 % de l'amplitude maximale transmise)
  - d'amplitude de signal transmis (normalisée à l'amplitude du signal délivré en entrée)
  - de niveau de diaphonie sur les interconnexions adjacentes à une interconnexion active lors de la présence de phénomènes de couplage (normalisé à l'amplitude du signal sur cette interconnexion active)

Ces caractéristiques des signaux sont intrinsèquement prises en compte dans les analyses des diagrammes de l'œil que nous avons effectuées mais elles ne sont pas quantifiées séparément. Il serait utile d'analyser ces autres critères de performance, en particulier la diaphonie, car les interprétations faites pourraient conduire à de nouvelles conclusions, et par conséquent à de nouveaux compromis sur les règles de dessin permettant d'optimiser les réseaux d'interconnexions des empilements 3D. A ce stade de nos travaux tous les outils de modélisation et de simulation sont en place pour mener ces études complémentaires.

- Etudier toutes les topologies possibles d'interconnexions représentait un travail trop large et nous n'avons privilégié que quelques architectures typiques et réalistes d'interconnexions d'empilements 3D connus. Les conclusions données dans nos « stratégies d'optimisation » sont par conséquent très (trop) génériques. Néanmoins, à chaque fois nous avons tenté de les chiffrer. Il est certain que cette quantification reste propre aux architectures d'interconnexions étudiées et que toute nouvelle architecture (changement de dimension, de longueur, d'isolant diélectrique...) conduira à des résultats différents sur les gains en termes de retard à 50 % ou sur les caractéristiques des diagrammes de l'œil. L'objectif de cette remarque est de préciser au lecteur que désormais nous

pouvons répondre à d'innombrables interrogations concernant les performances d'un réseau d'interconnexions pour des circuits 3D dès lors que nous pouvons définir son architecture et nous avons montré qu'il existait toujours une voie plus judicieuse que les autres pour optimiser les performances électriques en jouant sur le routage, le type d'orientation des puces ou la densité d'intégration.

Ces travaux ouvrent aussi la voie à une multitude d'opportunités. Une « bibliothèque », regroupant les caractéristiques électriques, des briques élémentaires d'interconnexion, présentes en intégration 3D, a été établie. Nous avons mis en place tous les éléments ainsi qu'un protocole complet, permettant l'étude en performance pour n'importe quelle application présente dans le domaine de l'intégration 3D.

Ainsi les performances de systèmes de type HDMI (Interface Multimédia Haute Définition), par exemple, composés de deux puces logiques et d'un BGA (Figure c-1), peuvent être analysées.

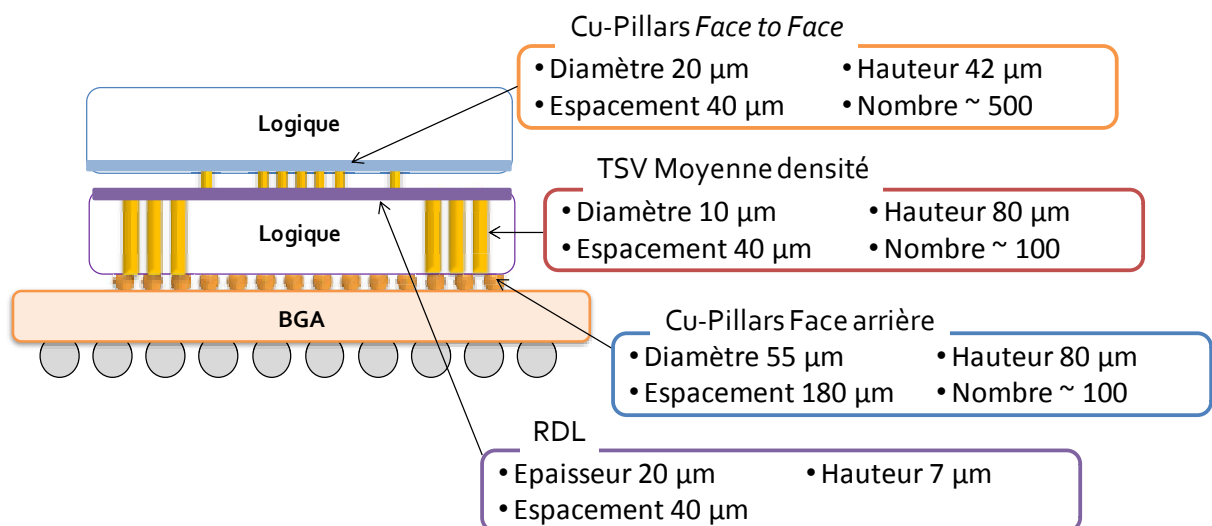


Figure c-1 : Exemple de structures HDMI en moyenne densité d'intégration

Une autre voie technologique de l'intégration 3D, très prometteuse, pourra aussi être expertisée. Il s'agit d'une solution d'empilement par interposer que nous pourrions qualifier « d'intelligent ». Cette approche consiste à rendre le plus fonctionnel possible l'interposer au sens où il ne réalisera pas uniquement son rôle « limité » de connexion électrique entre puces (ce qui est déjà une lourde tâche) mais de profiter au maximum de ce substrat de silicium pour intégrer des composants passifs ou des fonctionnalités, souvent encombrantes en termes de surface silicium. Dans cette optique de rendre plus intelligent l'interposer, les pistes que nous pouvons investiguer sont les suivantes :

- la conception d'inductances 2D planaires réalisées dans les couches épaisses de cuivre dédiées aux RDL. Le fait de bénéficier de ces couches épaisses devrait permettre de réaliser des inductances intégrées à fort facteur de qualité grâce à la minimisation des pertes métalliques, tout en offrant la possibilité de les réaliser sur plusieurs niveaux métalliques, donc avec un large choix de valeur d'inductance, et avec un haut niveau d'intégration [Ishida, 2010]
- la conception d'inductances 3D de type solénoïde réalisées dans le substrat même de l'interposer en tirant bénéfice des technologies de réalisation des TSV [RTI]. La Figure c-2 représente de manière schématique une telle architecture d'inductance 3D. Cette architecture d'inductance, quelque peu innovante, est très intéressante en termes de valeur d'inductance que l'on pourrait obtenir. Aujourd'hui les avis sont partagés sur cette possibilité, la critique majeure porte sur le noyau de ces solénoïdes... qui serait en silicium. Un facteur de qualité acceptable ne pourra être obtenu que dans le cas de silicium à très haute résistivité (HR)



Figure c-2 : Représentation schématique d'une inductance 3D dans un substrat silicium d'interposer

- la conception de capacités 3D intégrées dans le silicium de l'interposer et réalisées avec les technologies de fabrication des TSV [Roozeboom, 2011]. Dans ce cas on tire partie des techniques de gravure du silicium pour réaliser les vias afin de remplacer ces derniers par des structures capacitives Métal Isolant Métal (MIM) de topologie 3D, comme le rappelle la Figure c-3. On bénéficie alors de toute l'épaisseur du silicium pour maximiser les surfaces des électrodes et par conséquent les valeurs de capacités. L'encombrement et la surface de silicium utile sont réduits et il est possible d'atteindre des densités surfaciques de capacité élevées





Figure c-3 : Capacité MIM de type « TSV »

- la réalisation de TSV en structure coaxiale [Yin, 2011] [Khan, 2011] permettant la transmission de signaux à très haut débit, comme dans les applications Wide IO que nous avons étudiées, ou à très hautes fréquences. L'architecture coaxiale est bien connue pour ses qualités d'intégrité dans la transmission des signaux du fait de leur très large bande passante. Fabriquer des TSV possédant une architecture coaxiale a très rapidement été proposé malgré une faisabilité de fabrication technologique encore incertaine.

Comme le schématise la Figure c-4 l'intégration de ces différents composants passifs au sein d'un interposer offrirait des fonctionnalités supplémentaires à ce dernier, comme des fonctions de filtrage ou de découplage, et ainsi une valeur ajoutée certaine aux technologies 3D d'empilement de puces. Les performances de ces nouveaux composants passifs ne sont pas encore quantifiées et elles restent à être évaluées de manière in-situ, c'est-à-dire une fois ces composants réalisés dans leur environnement réel d'utilisation.

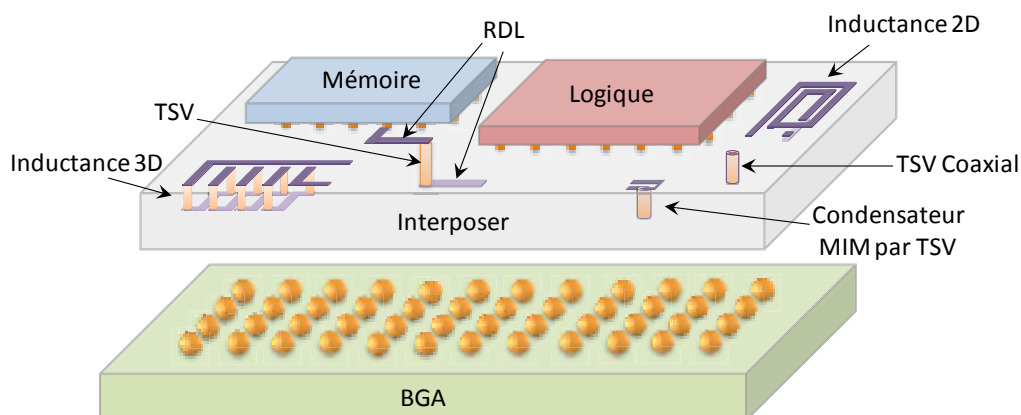


Figure c-4 : Schéma représentatif de la fonctionnalisation de l'interposer



# Références

---

- [Belleville, 2006] M. Belleville, N. Sillon, « *Intégration système : êtes-vous plutôt SiP ou SoC ?* », CEA Technologies, n°82 octobre, 2006.
- [Bermond, 2001] C. Bermond. « Caractérisation et modélisation des effets parasites liés aux interconnexions sur les performances électriques des circuits intégrés ultra-rapides », Thèse de Doctorat Université de Savoie, 2001.
- [Boeuf, 2007] F. Boeuf, M. Sellier, F. Payet, B. Borot, and T. Skotnicki, "Using Model for Assessment of Complementary Metal Oxide Semiconductor Technology and Roadmaps as a Pre-Simulation Program with Integrated Circuit Emphasis Model Generator for Early Technology and Circuit Simulation", Jpn. J. Appl. Phys, Vol. 47, pp. 3384-3389, 2008
- [Cadix-1, 2010] L. Cadix, M. Rousseau, C. Fuchs, P. Leduc *et al.* "Integration and frequency dependent electrical modeling of Through Silicon Vias (TSV) for high density 3DICs", Interconnect Technology Conference (IITC), 6-9 June 2010, San Francisco (CA).
- [Cao, 2003] Y. Cao, R. A. Groves, X. Huang, N. D. Zamdmer, J. Plouchart, R.A. Wachnik, T. King and C. Hu, "Frequency-independent equivalent-circuit model for on-chip spiral inductors", IEEE J. Solid-State Circuits, vol. 38, no. 3, pp. 419-426, Mar. 2003.
- [Chang, 2004] C. Chang *et al.*, "Trench filling characteristics of low stress TEOS/ozone oxide deposited by PECVD and SACVD", Microsystem Technologies, Vol. 10, n°2, pp 97-102, Janvier 2004.
- [Chen, 2000] Q. Chen, J. Zhao, "Via and return path discontinuity impact on high speed digital signal quality", Conf. On Electrical Performance of Electronic Packaging, pp. 215-218, 23-25 Octobre 2000.
- [Cheramy, 2009] S. Cheramy, D. Henry, A. Astier, J. Charbonnier *et al.* "3D integration process flow for set-top box application : description of technology and electrical results" EMPC, June 15th - 18th, 2009, Rimini, Italy.

- [De Rivaz, 2011] S. De Rivaz, « Développement d'outil de caractérisation et d'optimisation des performances électriques des réseaux d'interconnexion de circuits intégrés rapides sub-CMOS 65 nm et nouveaux concepts d'interconnexions fonctionnelles », Thèse de Doctorat Université de Grenoble, 2011.
- [Di Cioccio, 2005] L. Di Cioccio, B. Biasse, M. Kostrzeva *et al.* "Recent results on advanced molecular wafer bonding technology for 3D integration on silicon" *In ECS, The ninth International Symposium on Semiconductor Wafer Bonding*, pages 280-287, Quebec PQ, May 2005
- [Druais, 2010] G. Druais, P. Ancey, L.L. Chapelon, J. Charbonnier *et al* "TSV as an alternative to wire bonding for a wireless industrial product: another step towards 3D integration" *Proc of Electronic System-Integration Technology Conference (ESTC)*, 13-16 Sept. 2010.
- [Engen, 1979] G.F. Engen, C.A.Hoer. "Thru-Reflect-Line: An Improved Technique for Calibration the Dual Six-Port Automatic Network Analyser". *IEEE Mic. Theory & Tech. Vol. MTT-27*, n°12, pp.987-993, Déc. 1979
- [Fourneaud-1, 2011] L. Fourneaud, T. Lacrevez, J. Charbonier, A. Farcy, B. Flechet, "Innovative HF Extraction Procedure of the Characteristic Impedance for Embedded Planar Transmission Line on High Conductive Si Substrate", *Proc of Asia – Pacific Microwave Conference, Japan*, 7 – 10 December 2010.
- [Fourneaud-2, 2011] L. Fourneaud, T. Lacrevez, J. Charbonnier, C. Fuchs, A. Farcy, C. Bermond, E. Eid, J. Roullard, B.Flechet, "Extraction of Equivalent High Frequency Models for TSV and RDL Interconnects Embedded in Stacks of the 3D Integration Technology" *IEEE Signal Propagation on Interconnects*, 8-11 May 2011, Naples, Italy
- [Garrou-1, 2008] P. Garrou, C. Bower, P. Ramm. « Handbook of 3D Integration », *volume 1 – 1st Edition. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA*, 2008. P 210.
- [Garrou-2, 2008] P. Garrou, C. Bower, P. Ramm. « Handbook of 3D Integration », *volume 1 – 1st Edition. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA*, 2008. P 27.

- [Guillon, 2009] Y. Guillon, A.M. Dutron, "3D IC products using TSV for mobile phone applications: An industrial perspective", Proc of Microelectronics and Packaging Conference (EMPC), 15-18 June 2009
- [Ho, 2008] S.W. Ho, S.W. Yoon, Q. Zhou, K. Pasad, V. Kripesh, J.H. Lau, "High RF performance TSV silicon carrier for high frequency application", Proc. of Electronic Components and Technology Conference (ECTC), 27-30 may 2008
- [Hopkins, 2001] J. Hopkins et al. "Method and apparatus for etching a substrate", US patent, February 13, 2001
- [Ishida, 2011] K. Ishida, K. Takemura, K. Baba, M. Takamiya, T. Sakurai, "3D stacked buck converter with 15µm thick spiral inductor on silicon interposer for fine-grain power-supply voltage control in SiP's", IEEE International 3D Systems Integration Conference (3DIC), 16-18 Nov. 2010, Munich, pp 1-4
- [ITRS, 2005] Source internet:  
<http://www.itrs.net/Links/2005ITRS/ExecSum2005.pdf>
- [ITRS, 2009] Source Internet :  
[http://www.itrs.net/links/2009itrs/2009chapters\\_2009tables/2009\\_Interconnect.pdf](http://www.itrs.net/links/2009itrs/2009chapters_2009tables/2009_Interconnect.pdf)
- [JEDEC, 2011] Source internet :  
<http://www.jedec.org/standards-documents/technology-focus-areas>  
mobile memory : LPDDR2, LPDDR3, Wide IO, Memory MCP
- [Kajfez, 1986] D. Kajfez, W.P. Wheless, "Invariant Definitions of the Unloaded Q Factor" Microwave Theory and Techniques, Vol. 34, Issue 7, pp 840 – 841.
- [Katti, 2011] G. Katti, A. Mercha, J. Van Olmen, C. Huyghebaert *et al*, "3D stacked ICs using Cu TSVs and Die to Wafer Hybrid Collective bonding" Electron Devices Meeting (IEDM), 7-9 Dec. 2009.
- [Khan, 2011] N. H. Khan, S. M. Alam, S. Hassoun, "Power Delivery Design for 3-D ICs Using Different Through-Silicon Via (TSV) Technologies", IEEE Trans. On Very Large Scale Integration (VLSI) Systems, Vol. 19, No. 4, April 2011, pp 647-658

- [Kim, 2003] Kim, T. H.; Howlader, M. M. R.; Itoh, T.; Suga, T. "Room temperature Cu-Cu direct bonding using surface activated bonding method" *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, vol. 21, issue. no. 2, p. 449-453
- [Kim, 2006] B. Kim *et al.*, "Factors affecting copper filling process within high aspect ratio deep vias for 3D chip stacking", *Proceedings of the Electronic Components and Technology Conference (2006)* pp 838-843, San Diego (CA)
- [Kimmich, 2011] G. Kimmich, "3D FOR TABLETS AND SMARTPHONES" D43D, June 30 2011, Grenoble.
- [Klumpp, 2004] A. Klumpp, R. Merkel, P. Ramm, J. Weber and R. Wieland, "Vertical System Integration by Using Inter-Chip Vias and Solid-Liquid Interdiffusion Bonding" *Japanese Journal of Applied Physics* Vol. 43, No. 7A, 2004.
- [Kruppa, 1971] W. Kruppa, and K.F. Sodomsy. "An explicit solution for the scattering parameters of a linear two-port measured with an imperfect test set". *IEEE Trans. Microwave Théoriy Tech.*, Vol.19, pp.122-123, Jan.1971
- [Kwon, 2003] Y. Kwon, A. Jindal and J.J. McMahon, *et al.* "Dielectric glue wafer bonding for 3D ICs" *Technology and Reliability for Advanced Interconnects and Low-k Dielectrics-2003*, Materials Research Society, San Francisco, CA (2003), pp. 27-32.
- [Kwon, 2006] Y. Kwon, J. Seok, J.Q. Lu, *et al.* "Critical Adhesion Energy of Benzocyclobutene-Bonded Wafers" *Journal of Electrochemical. Soc.*, Vol. 153, Issue 4, pp. G347-G352 (2006).
- [Leduc, 2007] P. Leduc, F.D. Crécy, M. Fayolle, B. Charlet, T. Enot, M. Zussy, B. Jones, J.-C. Barbé, N. Kernevez, N. Sillon, S. Maitrejean, D. Louis, G. Passemard, "Challenges for 3D IC integration: bonding quality and thermal management", *Proceedings of IITC*, pp. 210 – 212 (2007).
- [Leduc, 2009] P. Leduc, M. Assous, L. DiCioccio, M. Zussy, T. Signamarcheix, A. Roman, M. Rousseau, S. Verrun, L. Bally, D. Bouchu, L. Cadix, A. Farcy, N. Sillon, "First integration of Cu TSV using die-to-wafer direct bonding and planarization", *Proceedings of IITC*, pp. 1 – 5 (2009).

- [Lee, 2006] K. Y. Lee, S. Mohammadi, P. K. Bhattacharya and L. Katehi, "Compact models based on transmission-line concept for integrated capacitors and inductors", *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 12, pp. 4141-4148, Dec. 2006.
- [Lèquepeys, 2011] Source internet : [http://www.techniques-ingenieur.fr/actualite/environnement-securite-energie-thematique\\_191/l-integration-3d-represente-un-enjeu-majeur-de-la-microelectronique-article\\_63036/](http://www.techniques-ingenieur.fr/actualite/environnement-securite-energie-thematique_191/l-integration-3d-represente-un-enjeu-majeur-de-la-microelectronique-article_63036/)
- [Lo, 2007] W.-C. Lo, S.-M. Chang, Y.-H. Chen, J.-D. Ko, T.-Y. Kuo, H.-H. Chang and Y.-C. Shih, "3D chip-to-chip stacking with through silicon interconnects", *International Symposium on VLSI-TSA*, 13, pp. 72-73. (2007).
- [Marchal, 2011] P. Marchal *et al*, "3D Technology Roadmap and Status" Interconnect Technology Conference (IITC), 9-12 May 2011, Dresden, Germany.
- [Meindl, 2001] J.D. Mendl, *et al*, "Interconnecting device opportunities for gigascale integration (GSI)", *Proc. of IEDM (Electron Devices Meeting, 2001. IEDM Technical Digest. International)*, Washington, Dc, December 2-5, 2001.
- [Meindl, 2002] D. Meindle, J.A. Davis, *et al*, "Interconnect opportunities for gigascale integration" *IBM Journal of Research and Development*, mars 2002.
- [Moore, 1965] G.E. Moore, "Cramming more components onto integrated circuits", *Electronics* 38 - 8 April 19, 1965.
- [Moriceau, 2003] H. Moriceau, O. Rayssac, B. Aspar, and B. Ghyselen. "The bondig energy control : an original way to debondable substrates". In *ECS, The seventh International Syposium on Semiconductor Wafer Bonding Science, Technology, and Applications*, pages 49-56, Paris, May 2003.
- [Murphy, 2005] O.H. Murphy, K. G. McCarthy, C. J. P. Delabie, A. C. Murphy, and P. J. Murphy, "Design of multiple-metal stacked inductors incorporating an extended physical model", *IEEE Trans. Microwave Theory Tech.*, vol. 53, no. 6, pp. 2063-2072, June 2005.

- [Nguyen, 2008] L. Nguyen Tran, D. Pasquet, E. Bourdel, S. Quintanel, "CAD-Oriented Model of a Coplanar Line on a Silicon Substrate Including Eddy Current Effects and Skin Effect", IEEE Transactions on Microwave Theory and Techniques, Vol.56, n°3, pp. 663-670, March 2008.
- [Nguyen, 2009] L. Nguyen Tran, « Caractérisations et modélisation d'interconnexions et d'inductances en technologies BiCMOS. Application à l'amplificateur faible bruit », Thèse de Doctorat Université de Cergy-Pontoise, 2009.
- [Nimura, 2011] M. Nimura, J. Mizuno, K. Sakuma, and S. Shoji, "Solder/Adhesive Bonding Using Simple Planarization Technique for 3D Integration" Proc 61th Electronic Components and Technology Conference, Lake Buena Vista, Florida, May 31- June 3, 2011.
- [Pozar, 2005] D.M. Pozar, « Microwave Engineering », Third edition, Wiley, 2005.
- [Pozder, 2008] S. Pozder, A. Jain, R. Chatterjee, Z. Huang, R.E. Jones, E. Acosta, B. Marlin, G. Hillmann, M. Sobczak, G. Krein, H. Kostner and S. Pargfrieder, "3D Die-to-wafer Cu/Sn Microconnects Formed Simultaneously with an Adhesive Dielectric Bond Using Thermal Compression Bonding", *Proceedings of IITC*, pp. 46 – 48 (2008).
- [Roozeboom, 2008] F. Roozeboom, W. Dekkers, Y. Lamy, J. Klootwijk, E. Grunsven, H. Kim "System in package integration of passives using 3D through-silicon vias." Solid State Technology, May 2008, pp 38-47
- [Roullard-1, 2011] J. Roullard, S. Capraro, T. Lacrevez, M. Gallitre, C. Bermond, A. Farcy, B. fléchet, "Characteristic impedance extraction of embedded and integrated interconnects", European Journal of Physics - applied Physics, March 2011, vol. 53, pp.33605
- [Roullard-2, 2009] J. Roullard, S. Capraro, T. Lacrevez, C. Bermond, A. Farcy, B. Fléchet, « Extraction de l'impédance caractéristique d'interconnexions », 16<sup>ème</sup> Journées Nationales Microondes, mai 2009, Grenoble
- [Rousseau, 2009] M. Rousseau, « Impact des technologies d'intégration 3D sur les performances des composants CMOS, Thèse de doctorat Université de Toulouse, 2009
- [RTI] Source internet : 3D Integration technology, RTI International, [www.rti.org/microsystem](http://www.rti.org/microsystem)



- [Sakurai, 1983] T. Sakurai, K. Tamaru, "Simple formulas for two- and three-dimensional capacitances" *Electron Devices, IEEE Transactions*, Vol. 30, Issue 2, pp 183-185, Feb 1983.
- [SIA, 2001] Semiconductor International Association, "International Technology Roadmap for Semiconductors", 2001 Edition. <http://public.itrs.net>.
- [Sillon, 2008] N.Sillon, A.Astier, H.Boutry, L. Di Cioccio, D.Henry, P. Leduc, "Enabling technologies for 3D Integration: From packaging miniaturization to advanced stacked ICs", *Proc. of International Electron Devices Meeting (IEDM)*, San Francisco, pp.1-4, 2008
- [Stengl, 1989] R. Stengl, T. Tan and U. Gösele, "A model of silicon wafer bonding process", *Japanese Journal of Applied Physics*, Vol.28, Issue 10, pp. 1735, 1989.
- [Tai, 2007] C. M. Tai and C. N. Liao, "A physical model of solenoid inductors on silicon substrates", *IEEE Trans. Microwave Theory Tech*, vol. 55, no. 12, pp. 2579-2585, Dec. 2007.
- [Taibi, 2010] R. Taibi, L. Di Cioccio, C. Chappaz et al "Full characterization of Cu/Cu direct bonding for 3D integration", *Proc of Electronic Components and Technology Conference (ECTC)*, Las Vegas, pp 219 – 225, 1-4 June 2010.
- [Takahashi, 2003] K. Takahashi, M. Umemotoa, N. Tanakaa, K. Tanidaa, Y. Nemotoa, Y. Tomitaa, M. Tagoa and M. Bonkoharab, "Ultra-high-density interconnection technology of three-dimensional packaging" *Microelectronics Reliability*, Vol. 43, Issue 8, Pages 1267-1279, August 2003.
- [Tezcan, 2006] D.S. Tezcan *et al*, "Development of vertical and tapered via etch for 3D through wafer interconnect technology", *Proceedings of the Electronics Packaging Technology Conference*, pp 22-28, 6-8 December 2006.
- [Turner, 2002] K.T. Turner and S.M. Spearing, "Modeling of direct wafer bonding : effect of wafer bow and etch patterns", *Journal of Applied Physics*, Vol.92, Issue 12, pp. 7658, 1989.
- [Twordzylo, 1998] W.Twordzylo, W. Cecot, J.T. Oden and C.H. Yew. "New asperity-based models of contact friction" *Wear*, 220, 113-140.

- [Vandamme, 2001] E. Vandamme, « Improved three-step de-embedding method to accurately account for the influence of pad parasitic in silicon on wafer RF test-structures » IEEE Trans. on Electron Devices, 48, 737 2001
- [Wang, 2003] Q. Wang, N. Hosoda, T. Itoh and T. Suga, "Reliability of Au bump-Cu direct interconnections fabricated by means of surface activated bonding method" *Microelectronics Reliability*, Vol. 43, Issue 5, Pages 751-756, May 2003.
- [Williams, 1995] D.F. Williams, J.B. Schappacher, "Line-Reflect-Match Calibrations with Nonideal Microstrip Standards", 46th ARFTG Conference Digest, pp 35-38, Nov. 30-Dec. 1, 1995.
- [Winkel-1, 1996] T.M. Winkel *et al.* "An accurate determination of the characteristics impedance of lossy lines on chips based on high frequency S-parameter measurements", IEEE Multi-chip module conference, pp. 190-195, Février 1996.
- [Winkel-2, 1996] T.M. Winkel *et al.* "An on-wafer deembedding procedure for devices under measurements with error-networks containing arbitrary line lengths", 47th Automatic Radio Frequency Techniques Group, pp. 102-111, June 1996.
- [Yin, 2011] W.Y. YIN, X.P. Wang, W.S. Zhao, X.L. Xu, "Frequency- and Temperature-Dependent Modeling of Coaxial Through-Silicon Vias for 3-D lcs", IEEE Transactions on Electron Devices, Volume: 58 Issue: 10, Oct. 2011, pp 3358 - 3368
- [Yole, 2009] Yole Développement, "Via First vs.Via Last? The first study to understand 3D integration scenarios", Final Report -December 2009.
- [Zhang, 2005] S.X. Zhang *et al.*, "Characterization of copper-to-silicon diffusion for the application of 3D packaging with through-silicon vias", Proceedings of the Electronics Packaging Technology Conference, 2005.

# Annexes

---

ANNEXE 1 : LE CALIBRAGE .....	198
ANNEXE 2 : EXTRACTION DE L'IMPEDANCE CARACTERISTIQUE DE LA LIGNE DE TRANSMISSION SOUS TEST PAR LA METHODE DE D.F. WILLIAMS.....	202
ANNEXE 3 : SIMPLIFICATION DU COEFFICIENT XP .....	206
ANNEXE 4 : CALCUL DE LA MATRICE $[A_{AP}]$ .....	208
ANNEXE 5 : CALCUL DE LA MATRICE $[A_{AL}]$ .....	209

## Annexe 1 : Le Calibrage

Le calibrage et le de-embedding sont des techniques complémentaires qui ont pour objectif d'extraire les grandeurs associées à notre DST seul, en corrigeant les erreurs engendrées par le contexte de mesure. Les erreurs, pouvant être corrigées par un calibrage adéquate, résultent de causes diverses : l'analyseur vectoriel de réseau lui-même, les sondes de mesures hyperfréquences, les câbles assurant la liaison entre les sondes et l'analyseur et les divers connecteurs introduits entre les sondes et l'analyseur.

D'une manière générale le calibrage consiste à repositionner les plans de référence de la mesure au plus proche du DST. Pour cela, on réalise des mesures complémentaires sur des dispositifs particuliers (ces derniers remplaçant alors le DST), appelés cellule de de-embedding (CDE), dont on connaît le comportement en haute fréquence.

### I.1 Les techniques de calibrage

Il existe plusieurs techniques de calibrage permettant de corriger les erreurs liées à l'ensemble de mesure analyseur de réseau/connectique. Chacun de ces calibrages fait appel à un type de CDE. Bien qu'implicitement une technique de calibrage se trouve généralement associée à un modèle d'erreur, elle peut alors se définir simplement par une spécification des CDE utilisées. On distingue entre autres les calibrages suivants [Bermond, 2001] :

Le calibrage SOLT dans lequel on utilise les standards SHORT, OPEN, LOAD et THRU (Figure 1).

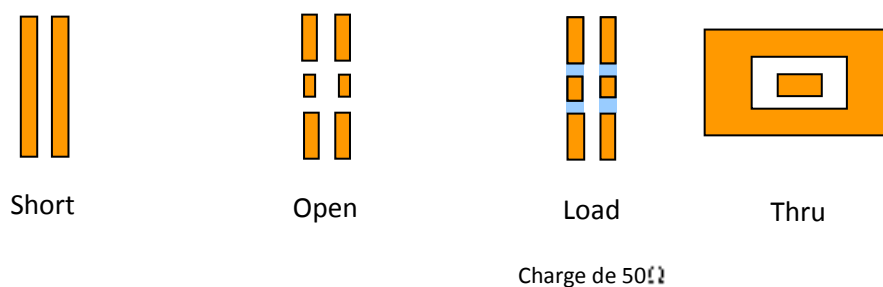


Figure 1 : Détail des CDE du calibrage OSTL

Le calibrage TRL (plus précisément auto-calibrage TRL) qui est basé sur les standards THRU, REFLECT, et LINE (Figure 2).

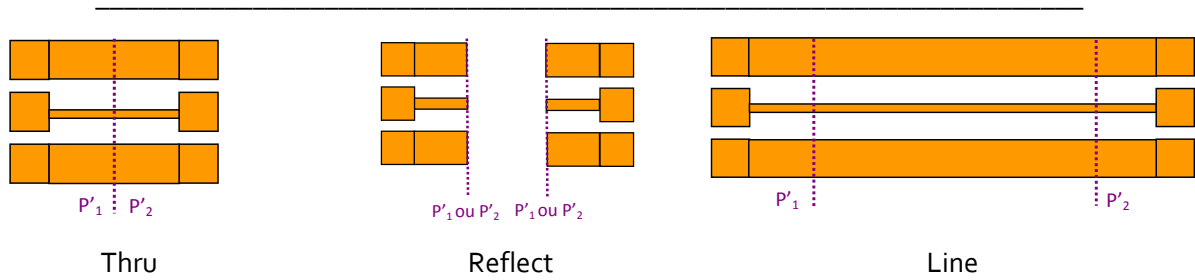


Figure 2 : Détail des CDE du calibrage TRL

En pratique, les VNA intègrent des procédures automatisées grâce auxquelles l'opérateur pourra sélectionner la technique de calibrage appropriée à sa problématique. L'opérateur devra alors présenter les standards pour mesure dès que le proposera le VNA. Ce dernier se chargera ensuite de calculer les termes d'erreur du modèle puis au final de passer en mode calibré.

### I.1.1 Le calibrage SOLT ou OSTL [Kruppa, 1971]

Dans la méthode SOLT, les dispositifs suivants seront mesurés:

Un court circuit associé au standard SHORT

Un circuit ouvert (OPEN)

Une charge adaptée (LOAD) dont l'impédance est généralement de 50 Ohms

Un tronçon de ligne (THRU) pour la connexion des deux ports de test dont l'impédance caractéristique est généralement de 50  $\Omega$ .

Dans le cas d'une ligne de transmission on obtient la configuration exposée Figure 3 :

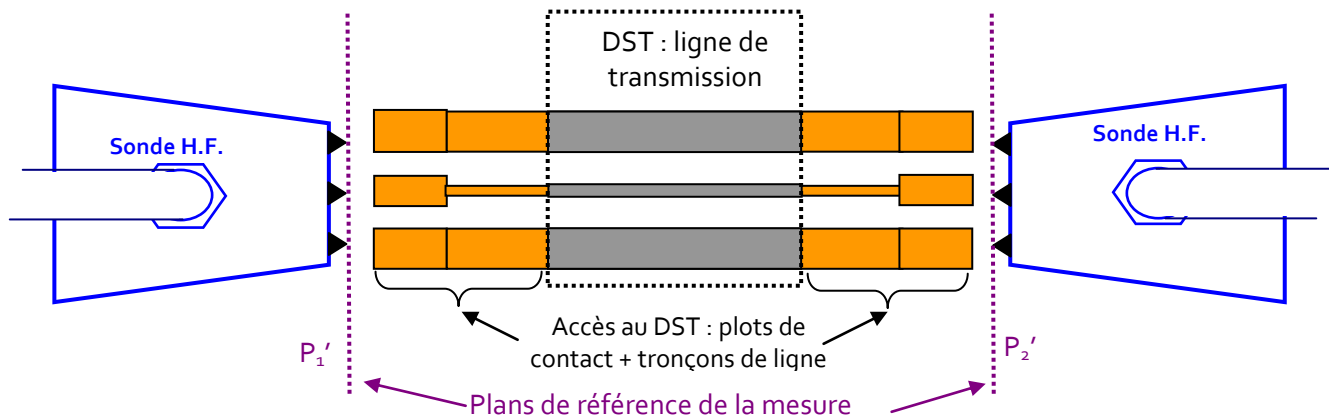


Figure 3 : Plans de référence après un calibrage OSTL situé au niveau des pointes des sondes HF

Ce type de calibrage nous permet en définitif de placer les plans de référence de la mesure au niveau des points de contacts des sondes de mesures (Figure 3), et d'être référencé à une impédance connue de 50  $\Omega$ . On notera qu'une fois les sondes H.F posées sur les plots de contacts, compte tenu de

la position des plans de référence de la mesure, feront partie de la mesure aussi bien la zone hachurée que les autres (Figure 4).

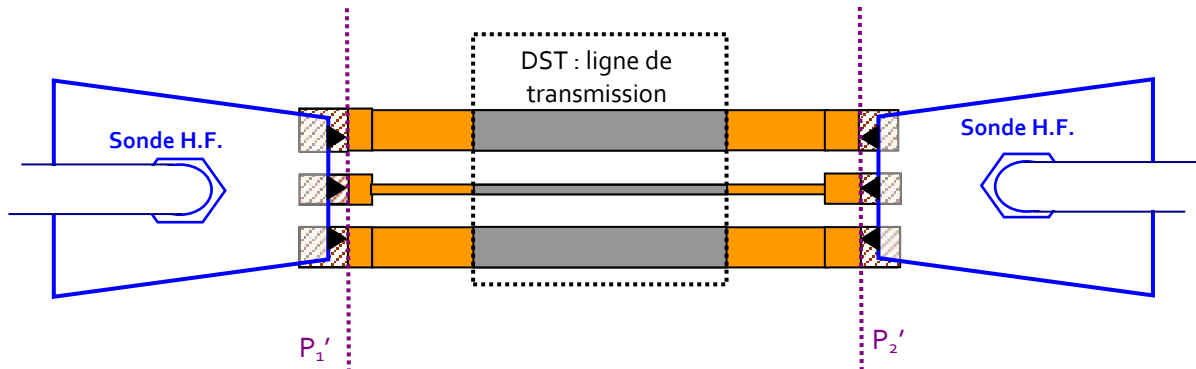


Figure 4 : Plans de référence après un calibrage OSTL ; la partie mesurée sera aussi bien la zone hachurée que les autres

Une mesure de paramètres  $S$  à la suite d'un calibrage OSTL implique que ces derniers sont donnés référencés à 50  $\Omega$ . On remarquera que la mesure d'un même dispositif donnera des paramètres  $S$  différents si l'impédance de référence associée au calibrage est différente. On notera que les plans de référence étant situés aux niveaux des sondes de mesures une étape de de-embedding sera nécessaire pour retirer de la mesure l'impact des plots de contacts.

### I.1.2 Le calibrage TRL [Engen, 1979].

Cette technique s'applique à partir de la mesure de trois dispositifs :

- Une ligne de transmission, associée à la CDE THRU, qui connecte les deux ports de test du VNA.
- Une ligne de transmission (CDE LINE) identique en tous points à celle associée à la CDE THRU mais dont la longueur est excédante.
- Une CDE permettant de fermer successivement chacun des accès 1 et 2 par une même charge donnant lieu à un coefficient de réflexion très élevé. (CDE REFLECT)

Dans le cas d'une ligne de transmission, on obtient la configuration exposée Figure 5 :

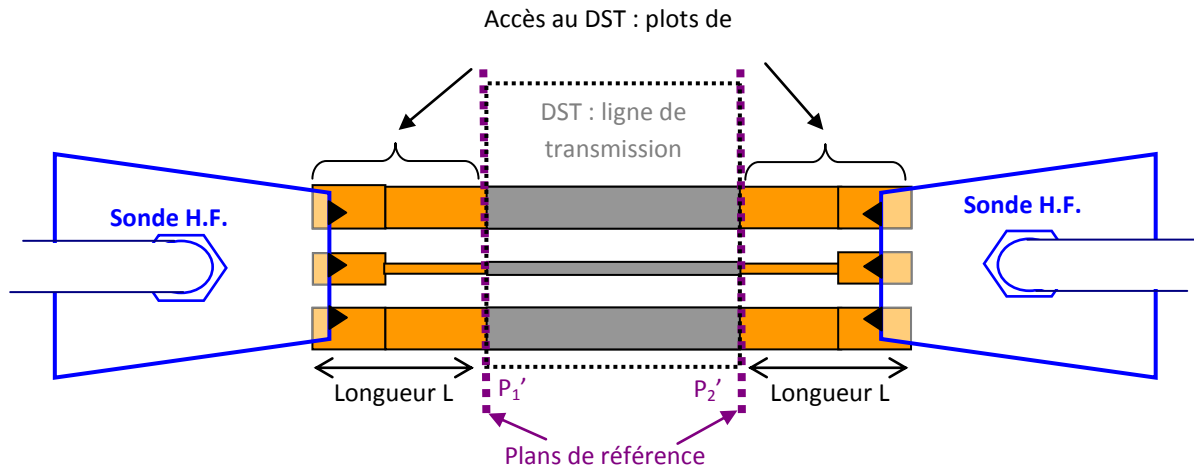


Figure 5 : Plan de référence après un calibrage TRL situé aux bornes du DST

Ce type de calibrage présente l'avantage de pouvoir placer les plans de référence de la mesure aux bornes du DST : ce dernier devant impérativement être accessible par 2 tronçons de circuits identiques respectivement à la moitié gauche et droite de la CDE THRU (Figure 6).

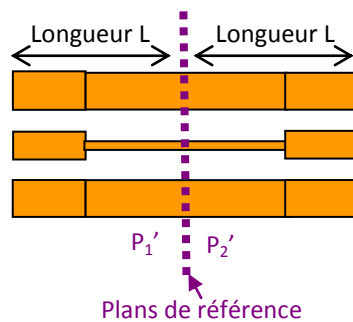


Figure 6 : CDE THRU et plans de référence de la mesure situés en son centre

L'impédance caractéristique de la ligne de transmission du standard LINE constituera l'impédance de référence des paramètres de la matrice [S]. Cette impédance est, dans le contexte de la microélectronique, inconnue dans la plupart des cas et est souvent la grandeur précisément recherchée.

## Annexe 2 : Extraction de l'impédance caractéristique de la ligne de transmission sous test par la méthode de D.F. Williams

Il s'agit maintenant de montrer la démarche pour calculer les paramètres de la matrice [Ts].

### I.2 Détermination des paramètres d'erreur de la transition plot ligne

Dans cette annexe nous proposons une méthode de calcul des paramètres de la matrice de transfert T de la transition plot ligne. Ces paramètres sont calculés à l'aide des paramètres S de la transition plot ligne. La détermination de ces paramètres S sera effectuée à l'aide d'une méthode basée sur l'approche du calibrage TRL.

La transition plot ligne peut se modéliser sous la forme d'une matrice S en entrée et en sortie du dispositif sous test. Ainsi il s'agit de calculer les paramètres de ces quadripôles d'entrée et de sortie (Figure 7).

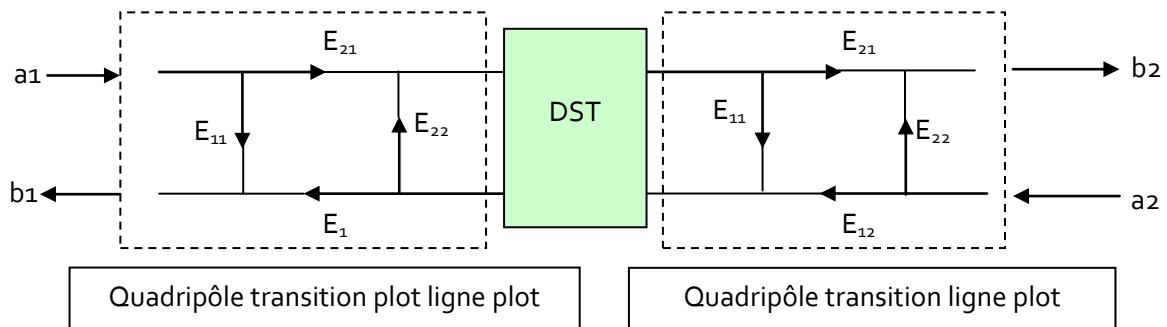


Figure 7 : Graphe de fluence modélisant la transition ligne plot en entrée et en sortie du standard de mesure sous forme de paramètres S

La matrice de transfert du quadripôle plot ligne entrée sera notée [Ta]. La matrice de transfert du quadripôle ligne plot en sortie sera notée [Tb].

La procédure de calibrage TRL comprend trois étapes de mesure : mesure de la CDE THRU, puis de la CDE LINE et enfin de la CDE REFLECT.



### I.2.1 Etape 1 et 2 : Mesure des CDE THRU et LINE

La CDE THRU est définie comme étant la connexion directe des deux quadripôles transition. La matrice S de la CDE THRU est donc par définition :

$$\mathbf{S}_t = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ et pour la matrice de transfert } \mathbf{T}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

La CDE LINE est définie comme étant la connexion entre les deux quadripôles transition d'un tronçon de ligne de transmission de longueur l et de structure identique à celle des accès au dispositif de test.

$$\mathbf{S}_l = \begin{bmatrix} 0 & e^{-\gamma l} \\ e^{-\gamma l} & 0 \end{bmatrix} \text{ et pour la matrice de transfert } \mathbf{T}_l = \begin{bmatrix} e^{-\gamma l} & 0 \\ 0 & e^{-\gamma l} \end{bmatrix}$$

Où  $\gamma$  est l'exposant de propagation de la ligne insérée.

La mesure des CDE THRU et LINE donnent respectivement les matrices :

$$[\mathbf{T}_{mt}] = [\mathbf{T}_a] [\mathbf{T}_t] [\mathbf{T}_b]$$

$$[\mathbf{T}_{ml}] = [\mathbf{T}_a] [\mathbf{T}_l] [\mathbf{T}_b]$$

On élimine  $[\mathbf{T}_b]$  par l'opération matricielle :

$$[\mathbf{T}_{ml}] [\mathbf{T}_{mt}]^{-1} = [\mathbf{T}_a] [\mathbf{T}_l] [\mathbf{T}_t]^{-1} [\mathbf{T}_a]^{-1}$$

On élimine  $[\mathbf{T}_a]$  en inversant la première relation en la multipliant par la seconde :

$$[\mathbf{T}_{mt}]^{-1} [\mathbf{T}_{ml}] = [\mathbf{T}_b]^{-1} [\mathbf{T}_t]^{-1} [\mathbf{T}_l] [\mathbf{T}_b]$$

$$\text{Or par définition } \mathbf{T}_{ml} \mathbf{T}_{mt}^{-1} = \begin{bmatrix} e^{-\gamma l} & 0 \\ 0 & e^{-\gamma l} \end{bmatrix}$$

Ces expressions permettent de calculer l'exposant car les matrices  $[\mathbf{T}_a]$  et  $[\mathbf{T}_b]^{-1}$  permettent de diagonaliser respectivement les produits  $[\mathbf{T}_l] [\mathbf{T}_t]^{-1}$  et  $[\mathbf{T}_t]^{-1} [\mathbf{T}_l]$ . On obtient les paramètres suivants :

$$\frac{T_{11}a}{T_{21}a} - E_{11}a = \frac{E_{12}aE_{21}a}{E_{22}a} \qquad \frac{T_{12}a}{T_{22}a} - E_{11}a$$

Et

$$\frac{T_{11}b}{T_{12}b} - E_{22}b \mid \frac{E_{12}bE_{21}b}{E_{11}b} \quad \frac{T_{21}b}{T_{22}b} - E_{22}b$$

A l'issue de ces deux étapes de mesure, les termes  $E_{11}a$  et  $E_{22}b$  sont connus ainsi que les rapports :

$$\frac{E_{12}aE_{21}a}{E_{22}a} \text{ et } \frac{E_{12}bE_{21}b}{E_{11}b}. \text{ L'exposant de propagation est également connu.}$$

### I.2.2 Etape 3 : Mesure de la CDE REFLECT

La CDE REFLECT est défini comme la terminaison par un même dispositif de fort coefficient de réflexion sur chacun des ports. La matrice  $S$  s'écrit alors :

$$\mathbf{\Gamma}_r = \begin{bmatrix} \Gamma_r & 0 \\ 0 & \Gamma_r \end{bmatrix}$$

La mesure du standard REFLECT donne respectivement  $\mathbf{\Gamma}_1^*$  sur le port 1 et  $\mathbf{\Gamma}_2^*$  sur le port 2.

$$\Gamma_1 = E_{11}a \mid \frac{E_{12}aE_{21}a}{\frac{1}{\Gamma_r} E_{22}a} \quad \Gamma_2 = E_{22}b \mid \frac{E_{12}bE_{21}b}{\frac{1}{\Gamma_r} E_{11}b}$$

En éliminant  $\mathbf{\Gamma}_r^{-1}$  dans les équations précédentes et en posant :

$$X = \frac{E_{12}aE_{21}a}{E_{22}a} \quad Y = \frac{E_{12}bE_{21}b}{E_{11}b}$$

$$A = \mathbf{\Gamma}_1^* - E_{11}a$$

$$B = \mathbf{\Gamma}_2^* - E_{22}b$$

$$C = S_{11mt} - E_{11}a$$

avec  $S_{11mt}$  mesure du coefficient de réflexion de la CDE THRU.

Le calcul de  $E_{11}b$  est effectué par :

$$E_{11}b = \left[ \left( 1 + \frac{X}{A} \right) \left( 1 + \frac{Y}{B} \right) \left( 1 + \frac{X}{C} \right) \right]^{1/2}$$

puis celui de  $E_{22}a$  par :

$$E_{22}a = E_{11}b \left( 1 + \frac{Y}{B} \right) \left( 1 + \frac{X}{A} \right)^{-1}$$

Les produits suivants sont alors déduits :

$$E_{12}aE_{21}a = X E_{22}a$$

$$E_{12}bE_{21}b = Y E_{11}b$$

Le choix du signe lors du calcul de  $E_{11}b$  est obtenu en calculant le coefficient  $\Gamma_r$ . Le mauvais choix du signe dans l'équation conduit à un saut de phase de  $180^\circ$  de la phase de  $\Gamma_r$ .

A l'issue de cette troisième étape. Les paramètres suivants sont déterminés :  $E_{22}a$ ,  $E_{11}b$ . De plus les produits suivants sont également déterminés :  $E_{12}aE_{21}a$  et  $E_{12}bE_{21}b$

Le calcul de la matrice de transfert de la transition plot ligne nécessite de connaître chacun des éléments de la matrice de réparation de la transition. Compte tenu de la réciprocité de la transition nous obtenons  $E_{12}a = E_{21}a$ . Dans ces conditions on fait l'hypothèse suivante :  $E_{12}a = (E_{12}aE_{21}a)^{-1/2}$ .

Les quatre paramètres S de la transition plot ligne sont donc connus. Le calcul de la matrice de transfert Ts et de l'impédance caractéristique sont maintenant possibles.

### Annexe 3 : Simplification du coefficient XP

Le coefficient XP est exprimé en fonction des termes extra diagonaux de la matrice ABCD correspondant à la matrice globale  $[T_{Thp}] = [T_A][T_{L2}] \cdot [T_{L1}]^{-1} [T_A]^{-1}$

Matrice  $[T_A]$  :

$$[T_A] = \begin{bmatrix} T_{A11} & T_{A12} \\ T_{A21} & T_{A22} \end{bmatrix}$$

Inverse de la matrice  $[T_A]^{-1}$  :

$$[T_A]^{-1} = \begin{bmatrix} \frac{T_{A22}}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} & -\frac{T_{A12}}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} \\ -\frac{T_{A21}}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} & \frac{T_{A11}}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} \end{bmatrix}$$

Matrice du tronçon de ligne  $[T_{L2}] \cdot [T_{L1}]^{-1}$  :

$$[T_{L2}] \cdot [T_{L1}]^{-1} = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}$$

Avec

$$X = e^{\gamma_p (L_2 - L_1)}$$

$$Y = e^{-\gamma_p (L_2 - L_1)}$$

Matrice globale  $[T_{Thp}] = [T_A][T_{L2}] \cdot [T_{L1}]^{-1} [T_A]^{-1}$

$$[T_{Thp}] = \begin{bmatrix} \frac{T_{A11} \cdot T_{A22} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} - \frac{T_{A12} \cdot T_{A21} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} & -\frac{T_{A11} \cdot T_{A12} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} + \frac{T_{A11} \cdot T_{A12} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} \\ \frac{T_{A21} \cdot T_{A22} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} - \frac{T_{A21} \cdot T_{A22} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} & -\frac{T_{A12} \cdot T_{A21} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} + \frac{T_{A11} \cdot T_{A22} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} \end{bmatrix}$$

Termes extra diagonaux de la matrice ABCD correspondant à la matrice globale  $[T_{Thp}]$  :

$$A_{thp12} = \frac{(T_{Thp11} + T_{Thp21} - T_{Thp12} - T_{Thp22})}{2}$$

$$A_{thp12} = \frac{1}{2} \left( \frac{T_{A11} \cdot T_{A12} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} + \frac{T_{A12} \cdot T_{A21} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} + \frac{T_{A11} \cdot T_{A22} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} \right. \\ \left. + \frac{T_{A21} \cdot T_{A22} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} - \frac{T_{A11} \cdot T_{A12} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} - \frac{T_{A12} \cdot T_{A21} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} \right. \\ \left. - \frac{T_{A11} \cdot T_{A22} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} - \frac{T_{A21} \cdot T_{A22} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} \right)$$

$$A_{thp21} = \frac{(T_{Thp11} + T_{Thp12} - T_{Thp21} - T_{Thp22})}{2}$$

$$A_{thp21} = \frac{1}{2} \left( -\frac{T_{A11} \cdot T_{A12} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} + \frac{T_{A12} \cdot T_{A21} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} + \frac{T_{A11} \cdot T_{A22} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} \right. \\ \left. - \frac{T_{A21} \cdot T_{A22} \cdot X}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} + \frac{T_{A11} \cdot T_{A12} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} - \frac{T_{A12} \cdot T_{A21} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} \right. \\ \left. - \frac{T_{A11} \cdot T_{A22} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} + \frac{T_{A21} \cdot T_{A22} \cdot Y}{-T_{A12} \cdot T_{A21} + T_{A11} \cdot T_{A22}} \right)$$

Le coefficient  $XP^2 = \frac{A_{thp12}}{A_{thp21}}$

Or on sait que la matrice de transfère  $[T_A]$ , correspondant à la matrice ABCD  $[A_{AP}]$ , s'exprime de la façon suivante :

$$T_{A22} = (A_{AP11} + A_{AP12} + A_{AP21} + A_{AP22})/2$$

$$T_{A21} = (A_{AP11} - A_{AP12} + A_{AP21} - A_{AP22})/2$$

$$T_{A22} = (A_{AP11} + A_{AP12} - A_{AP21} - A_{AP22})/2$$

$$T_{A22} = (A_{AP11} + A_{AP12} - A_{AP21} - A_{AP22})/2$$

En remplaçant les termes  $T_A$  par leurs expressions en ABCD, le coefficient XP se simplifie de la manière suivante :

$$XP^2 = \frac{(A_{AP11} + A_{AP12}) \cdot (A_{AP11} - A_{AP22})}{(A_{AP12} - A_{AP21}) \cdot (A_{AP21} + A_{AP22})}$$

XP est exprimé en fonction des termes de la matrice  $[A_{AP}]$ .

### Annexe 4 : Calcul de la matrice $[A_{AP}]$

La matrice  $[A_{AP}]$  est composée des éléments suivants :  $[A_{AP}] = [A_{r1}] \cdot [A_0] \cdot [A_P] \cdot [A_{r2}]$

La matrice de normalisation d'entrée :

$$A_{r1} = \begin{bmatrix} \frac{1}{\sqrt{Z_0}} & 0 \\ 0 & \sqrt{Z_0} \end{bmatrix} \text{ avec } Z_0 = 50 \, \Omega, \text{ impédance au bout des pointes après un calibrage OSTL.}$$

La matrice de réflexion :

$$A_0 = \begin{bmatrix} 1 & 0 \\ \frac{\tanh(\gamma_p \cdot L_0)}{Z_p} & 1 \end{bmatrix}$$

La matrice de transmission :

$$A_P = \begin{bmatrix} \cosh(\gamma_p \cdot L_P) & Z_p \cdot \sinh(\gamma_p \cdot L_P) \\ \frac{\sinh(\gamma_p \cdot L_P)}{Z_p} & \cosh(\gamma_p \cdot L_P) \end{bmatrix}$$

La matrice de normalisation de sortie :

$$A_{r2} = \begin{bmatrix} \sqrt{Z_p} & 0 \\ 0 & \frac{1}{\sqrt{Z_p}} \end{bmatrix}$$

D'où la matrice  $[A_{AP}] = [A_{r1}] \cdot [A_0] \cdot [A_P] \cdot [A_{r2}]$  :

$$A_{AP} = \begin{bmatrix} \frac{\sqrt{Z_p} \cdot \cosh(\gamma_p \cdot L_P)}{\sqrt{Z_0}} & \frac{Z_p \cdot \sinh(\gamma_p \cdot L_P)}{\sqrt{Z_0} \sqrt{Z_p}} \\ \sqrt{Z_p} \left( \frac{\sqrt{Z_0} \cdot \sinh(\gamma_p \cdot L_P)}{Z_p} + \frac{\sqrt{Z_0} \cdot \cosh(\gamma_p \cdot L_P) \cdot \tanh(\gamma_p \cdot L_0)}{Z_p} \right) & \frac{\sqrt{Z_0} \cdot \cosh(\gamma_p \cdot L_P) + \frac{\sqrt{Z_0} \cdot Z_p \cdot \sinh(\gamma_p \cdot L_P) \cdot \tanh(\gamma_p \cdot L_0)}{Z_p}}{\sqrt{Z_p}} \end{bmatrix}$$

## Annexe 5 : Calcul de la matrice $[A_{AL}]$

La matrice  $[A_{AL}]$  est composée des éléments suivants :  $[A_{AL}] = [A_{r1}] \cdot [A_0] \cdot [A_P] \cdot [A_s] \cdot [A_{r2}]$

Les matrices  $[A_{r1}]$ ,  $[A_0]$  et  $[A_P]$  sont identiques à celles présentées dans l'annexe précédente.

Il reste à définir les matrices  $[A_s]$  et  $[A_{r2}]$ .

La matrice de transmission  $[A_s]$  :

$$A_s = \begin{bmatrix} \cosh(\gamma_s L_s) & Z_C \sinh(\gamma_s L_s) \\ \frac{\sinh(\gamma_s L_s)}{Z_C} & \cosh(\gamma_s L_s) \end{bmatrix}$$

La matrice de normalisation de sortie :

$$A_{r2} = \begin{bmatrix} \sqrt{Z_C} & 0 \\ 0 & \frac{1}{\sqrt{Z_C}} \end{bmatrix}$$

D'où la matrice  $[A_{AL}] = [A_{r1}] \cdot [A_0] \cdot [A_P] \cdot [A_s] \cdot [A_{r2}]$ :

$$A_{AL11} = \sqrt{Z_C} \left( \frac{\cosh(\gamma_P L_P) \cdot \cosh(\gamma_s L_s)}{\sqrt{Z_0}} + \frac{Z_P \sinh(\gamma_P L_P) \cdot \sinh(\gamma_s L_s)}{\sqrt{Z_0} Z_C} \right)$$

$$A_{AL12} = \frac{\frac{\cosh(\gamma_s L_s) \cdot Z_P \sinh(\gamma_P L_P)}{\sqrt{Z_0}} + \frac{\cosh(\gamma_P L_P) \cdot Z_C \sinh(\gamma_s L_s)}{\sqrt{Z_0}}}{\sqrt{Z_C}}$$

$$A_{AL21} = \sqrt{Z_C} \left( \cosh(\gamma_s L_s) \left( \frac{Z_0 \sinh(\gamma_P L_P)}{Z_P} + \frac{\sqrt{Z_0} \cosh(\gamma_P L_P) \tanh(\gamma_P L_0)}{Z_P} \right) \right. \\ \left. + \frac{1}{\sqrt{Z_C}} \left( \sinh(\gamma_s L_s) \left( Z_0 \cosh(\gamma_P L_P) + \frac{\sqrt{Z_0} Z_P \sinh(\gamma_P L_P) \tanh(\gamma_P L_0)}{Z_P} \right) \right) \right)$$

$$A_{AL22} = \frac{1}{\sqrt{Z_c}} \left( Z_c \cdot \text{Sinh}(\gamma_s, L_s) \left( \frac{\sqrt{Z_0} \cdot \text{Sinh}(\gamma_p, L_p)}{Z_p} + \frac{\sqrt{Z_0} \cdot \text{Cosh}(\gamma_p, L_p) \cdot \text{Tanh}(\gamma_p, L_0)}{Z_p} \right) \right. \\ \left. + \text{Cosh}(\gamma_s, L_s) \left( \sqrt{Z_0} \cdot \text{Cosh}(\gamma_p, L_p) + \frac{\sqrt{Z_0} \cdot Z_p \cdot \text{Sinh}(\gamma_p, L_p) \cdot \text{Tanh}(\gamma_p, L_0)}{Z_p} \right) \right)$$



# *Liste des publications*

---

## **Publications comme premier auteur : 11**

Revue internationale avec comité de lecture : 2

Conférences internationales avec comité de lecture : 5

Conférences nationales avec comité de lecture : 4

## **Publications comme co-auteur : 7**

Revue internationale avec comité de lecture : 1

Conférences internationales avec comité de lecture : 5

Conférences nationales avec comité de lecture : 1

## Publications comme premier auteur

**J. Roullard**, S. Capraro, T. Lacrevez, M. Gallitre, C. Bermond, A. Farcy, B. Fléchet « *Méthode d'extraction de l'impédance caractéristique d'interconnexions de circuits intégrés avancés et application à la caractérisation des isolants poreux ULK* » 6<sup>e</sup> Journées Franco-Maghrébines des Microondes et leurs Applications, 12-14 mars 2009, Agadir, Maroc

**J. Roullard**, S. Capraro, T. Lacrevez, C. Bermond, A. Farcy, B. Fléchet « *Extraction de l'impédance caractéristique d'interconnexions* » XVI<sup>e</sup> Journées Nationales Microondes, 27-29 mai 2009, Grenoble, France

**J. Roullard**, S. Capraro, T. Lacrevez, L. Cadix, E. Eid, A. Farcy, B. Flechet, "*Influence of 3D Integration on 2D Interconnections and 2D Self Inductors HF Properties*", IEEE Int. Conference on 3D System Integration (3D-IC), September 28-30, 2009, San Francisco, USA

**J. Roullard**, S. Capraro, E. Eid, L. Cadix, C. Bermond, T. Lacrevez, A. Farcy, B. Flechet "*Quality Factor and Frequency Bandwidth of 2D Self Inductors in 3D Integration Stack*", Material for Advanced Metallization, MAM 2010, 7-10 march 2010, Mechelen, Belgium

**J. Roullard**, S. Capraro, E. Eid, L. Cadix, T. Lacrevez, C. Bermond, A. Farcy, B. Fléchet, « *Influence des substrats de silicium sur les performances des inductances planaires 2D lors d'une intégration 3D* », 11<sup>e</sup> Journées Caractérisation Microondes et Matériaux, 31 mars - 2 avril 2010, Brest

**J. Roullard**, S. Capraro, T. Lacrevez, L. Cadix, E. Eid, A. Farcy, B. Flechet "*HF Performance Characterization and Prediction of 2D Redistribution Layer Interconnects in a 3D-Integrated Circuit Stack*" IEEE Signal Propagation on Interconnects, 9-12 May 2010, Hidelshheim, Germany

**J. Roullard**, S. Capraro, C. Bermond, A. Farcy, B. Fléchet, « *Caractérisation et optimisation des performances RF des interconnexions pour les architectures 3D de circuits intégrés* », 7<sup>e</sup> Journées Franco-Maghrébines des Microondes et Applications, 16-18 mars 2011, Tanger, Maroc

**J. Roullard**, S. Capraro, E. Eid, L. Cadix, C. Bermond, T. Lacrevez, A. Farcy, B. Flechet "Quality Factor and Frequency Bandwidth of 2D Self Inductors in 3D Integration Stack", Microelectronic Engineering, Elsevier, vol.88, issue 5, may 2011, pp. 734-738

**J. Roullard**, S. Capraro, T. Lacrevez, M. Gallitre, C. Bermond, A. Farcy, B. Fléchet "Characteristic Impedance Extraction of Embedded and Integrated Interconnects" European Physical Journal - Applied Physics, vol. 53, nb. 3, pp. 1-6, 2011

**J. Roullard**, S. Capraro, T. Lacrevez, C. Bermond, A. Farcy, J. Charbonnier, B. Flechet, "Electrical Characterization and Impact on Signal Integrity of New Basic Interconnection Elements Inside 3D Integrated Circuits", IEEE 61<sup>th</sup> Electronic Components and Technology Conference, June 1-3, 2011, Lake Buena Vista, Florida, USA

**J. Roullard**, A. Farcy, S. Capraro, T. Lacrevez, C. Bermond, J. Charbonnier, C. Fuchs, C. Ferrandon, P. Leduc, B. Flechet, "Evaluation of 3D Interconnect Routing and Stacking Strategies to Optimize High Speed Signal Transmission for Memory on Logic", IEEE 62<sup>nd</sup> Electronic Components and Technology Conference (ECTC), May 29 - June 1, 2012, San Diego, California, USA

## Publications comme co-auteur

L. Cadix, A. Farcy, C. Bermond, C. Fuchs, P. Leduc, M. Rousseau, M. Assous, A. Valentian, **J. Roullard**, E. Eid, N. Sillon, B. Fléchet, P. Ancey, "Modelling of Through Silicon Via RF Performance and Impact on Signal Transmission in 3D Integrated Circuits", IEEE Int. Conference on 3D System Integration (3D-IC), September 28-30, 2009, San Francisco, USA

E. Eid, T. Lacrevez, C. Bermond, S. de Rivaz, S. Capraro, **J. Roullard**, L. Cadix, B. Fléchet, A. Farcy, P. Ancey, F. Calmon, O. Valorge, P. Leduc, "Characterization and Modelling of Substrate Coupling Effects in 3D Integrated Circuit Stacking", Material for Advanced Metallization, MAM 2010, 7-10 march 2010, Mechelen, Belgium

E. Eid, T. Lacrevez, C. Bermond, S. de Rivaz, S. Capraro, **J. Roullard**, B. Fléchet, F. Calmon, C. Gontrand, A. Farcy, P. Ancey, « Effets de couplage RF par les substrats de silicium dans les empilements de circuits intégrés 3D », 11<sup>e</sup> Journées Caractérisation Microondes et Matériaux, 31 mars - 2 avril 2010, Brest

E. Eid, T. Lacrevez, C. Bermond, S. De Rivaz, S. Capraro, **J. Roullard**, L. Cadix, B. Fléchet, A. Farcy, P. Ancey, F. Calmon, O. Valorge, P. Leduc, "*Characterization and Modeling of RF Substrate Coupling Effects due to Vertical Interconnects in 3D Integrated Circuit Stacking* ", IEEE Signal Propagation on Interconnects, 9-12 May 2010, Hidelshiem, Germany

E. Eid, T. Lacrevez, C. Bermond, S. Capraro, **J. Roullard**, B. Fléchet L. Cadix, A. Farcy, P. Ancey, F. Calmon, O. Valorge, P. Leduc, "*Frequency and Time Domain Characterization of Substrate Coupling Effects in 3D Integration Stack*", 34<sup>th</sup> Int. Electronics Manufacturing Technology Conf., Nov. 30 - Dec. 2, 2010, Melaka, Malaysia

E. Eid, T. Lacrevez, C. Bermond, S. de Rivaz, S. Capraro, **J. Roullard**, L. Cadix, B. Fléchet, A. Farcy, P. Ancey, F. Calmon, O. Valorge, P. Leduc, "Characterization and Modelling of Substrate Coupling Effects in 3D Integrated Circuit Stacking", Microelectronic Engineering, Elsevier, vol.88, issue 5, may 2011, pp. 729-733

L. Fourneaud, T. Lacrevez, J. Charbonnier, C. Fuchs, A. Farcy, C. Bermond, E. Eid, **J. Roullard**, B. Flechet, "*Extraction of Equivalent High Frequency Models for TSV and RDL Interconnects Embedded in Stacks of the 3D Integration Technology*" IEEE Signal Propagation on Interconnects, 8-11 May 2011, Naples, Italy